

CAF - WORKING PAPER #2024/04

First version: July 05, 2024

This version: June 30, 2026

Money and gender: proxy respondents and the mismeasurement of income in surveys

Ricardo Estrada¹ | Daniela Goyheix² | María Lombardi³

¹CAF–development bank of Latin America and the Caribbean.

restrada@caf.com

²CAF–development bank of Latin America and the Caribbean.

daniela.goyheix_externo@caf.com

³Universidad Torcuato Di Tella.

mlombardi@utdt.edu

When sampled individuals are not found at home, many surveys rely on a proxy respondent: another knowledgeable household member. We study the effect of proxy reporting on the measurement of labor income in Mexico. For identification, we use the panel structure of the Mexican labor force survey and compare workers' reported earnings when they report them themselves to their reported earnings when another household member does the reporting. We find that male workers' monthly earnings are 7.2% lower when reported by a proxy, whereas female workers' earnings are only 1.8% lower. We provide evidence that this gender gap reflects both compositional effects, linked to men's higher earnings, and household informational asymmetries, partly driven by men's greater propensity to hide income from other household members. Finally, we study the implications of using proxy respondents and find that it can lead to an underestimation of the gender wage gap by 25%.

KEYWORDS

survey design, proxy reporting, labor income, household economics, gender wage gap

CAF - DOCUMENTO DE TRABAJO #2024/04

Primera versión: 5 de Julio, 2024

Esta versión: 30 de Junio, 2026

Dinero y género: informantes proxy y la medición incorrecta del ingreso en encuestas

Ricardo Estrada¹ | Daniela Goyheix² | María Lombardi³

¹CAF-banco de desarrollo de América Latina y el Caribe.
restrada@caf.com

²CAF-banco de desarrollo de América Latina y el Caribe.
daniela.goyheix_externo@caf.com

³Universidad Torcuato Di Tella.
mlombardi@utdt.edu

Cuando las personas seleccionadas para una muestra están ausentes, muchas encuestas recurren a un tercero que las represente — otro miembro del hogar que tenga conocimiento. En este trabajo estudiamos la diferencia entre el ingreso laboral informado por uno mismo (auto reportado) y el informado por un representante en México. Utilizamos la estructura de panel de la encuesta nacional de empleo y comparamos el ingreso de un mismo trabajador cuando lo reporta él mismo frente a cuando lo reporta otro miembro del hogar. Encontramos que los salarios mensuales de los hombres son 7,2 % más bajos cuando son reportados por un tercero, mientras que para las mujeres únicamente son 1,8 % menores. Proporcionamos evidencia de que la discrepancia en el ingreso reportado de los trabajadores masculinos se debe a una asimetría de información dentro del hogar, en parte debido a que los hombres ocultan ingresos a sus familiares. Finalmente, estudiamos las implicancias de usar respuestas de terceros en las encuestas y encontramos que esto puede producir una subestimación de la brecha salarial de género en un 25 %.

KEYWORDS

diseño de encuestas, informantes sustitutos, ingresos laborales, economía del hogar, brecha salarial de género

1 | INTRODUCTION

When sampled individuals are not found at home, many surveys rely on proxy respondents (other knowledgeable household members). For example, nearly half of interviews in the Current Population Survey (CPS) in the United States are done by proxies [U.S. Census Bureau \(2019\)](#). The use of proxy respondents helps to lower costs, but may introduce measurement error in reported variables, such as labor income. If classic, measurement error could lead to less precise estimates. Of greater concern is that measurement error could introduce bias if proxy respondents systematically under- or overestimate workers' earnings. This could happen, for example, if there is asymmetric information in the household due to individuals' desire to have greater discretionary power over their income [Deschênes et al. \(2020\)](#). By contrast, the use of proxy respondents could reduce bias if proxies are more likely to report workers' true earnings. This could be the case, for example, if individuals are less subject to social desirability bias when reporting other household members' earnings—as opposed to their own earnings.

Despite its relevance, there is limited understanding of how the use of proxy respondents affects reported earnings, and the related mechanisms and implications. This research gap is probably due to a methodological challenge, namely that for a given worker, we typically observe either self- or proxy-reported earnings. This is problematic because workers who report their own income and those who use a proxy respondent likely differ in unobservable characteristics correlated with income, thereby compromising the causal interpretation of a naive comparison between the reported earnings of these two types of workers. In this paper, we overcome this difficulty by using the panel structure of the Mexican labor force survey and the fact that some individuals (around half of our sample) respond to their own questionnaires in some survey waves, but not in others. This variation enables us to compare these workers' earnings when they report them themselves to their earnings when another household member does the reporting. Our identification assumption is that for a given worker, having a proxy respondent in a survey wave does not correlate with changes in actual (unobserved) labor income. We provide strong evidence in support of this assumption. That is, we show that our results are not driven by temporary

shocks linked to changes in jobs, working hours, bonuses, etc.

For the analysis, we use microdata from the Mexican labor force survey (ENOE as its Spanish acronym) for the 2005–2019 period. The ENOE has a rotating panel design in which sampled households are followed over five consecutive survey waves. Enumerators apply an employment questionnaire to all household members 15 years and older. If an individual is not at home at the time of the visit, another knowledgeable member of the household acts as a proxy. Crucially, the ENOE dataset identifies the informant. To mitigate concerns that potential income shocks might be confounded with proxy reporting, we restrict our sample to individuals who work full-time in a salaried position in all survey waves.

We perform the analysis separately for male and female workers. We begin by examining whether proxy reporting affects the likelihood that earnings are reported at all. Proxy respondents are less likely than workers themselves to report workers' earnings, with clear gender differences: proxy reports reduce the probability of observing earnings by 13.1 percentage points for male workers and 17.6 percentage points for female workers. For reference, approximately 90% of workers report their earnings when interviewed directly. Note that these differences in response rates do not pose a threat to the internal validity of our main research question, because we are interested in the effect of proxies on *reported* labor income.¹ We then turn to the effect of proxy respondents on reported earnings. We find a negative effect of proxy respondents on reported earnings, with substantial differences across genders. Men's monthly earnings are 7.2% lower when reported by a proxy, whereas the corresponding gap among female workers is only 1.8%.² Furthermore, we show that this average effect masks substantial heterogeneity in reported earnings by informant type for both genders.

Importantly, our results are robust to limiting our sample to workers who are less likely to have experienced income fluctuations throughout survey waves (salaried workers

¹A different story would be if we were studying the effect of proxies on *actual* earnings (and proxying these with reported earnings). In that case, the effect of the treatment on the probability of observing the outcome of interest could introduce selective attrition and lead to a biased estimate.

²Income misreporting is an old concern in survey design and there are multiple papers devoted to estimating, for example, the extent to which income is underreported in surveys compared to in administrative sources such as tax records [Mellow and Sider \(1983\)](#); [Kapteyn and Ypma \(2007\)](#); [Abowd and Stinson \(2013\)](#); [Meyer et al. \(2015\)](#). Rather than estimating overall misreporting, the purpose of our paper is to understand the marginal effect of using proxy reporting on the measurement of income. In other words, our results can be interpreted as an effect on top of any underlying tendency of, for example, rich individuals to underreport their income.

that do not receive bonuses, who remain at the same firm, who work regular hours, etc.), indicating that our findings stem from reporting discrepancies rather than from changes in actual earnings.

Turning to mechanisms, our findings indicate that the gender gap in the probability of reporting workers' labor income is mainly driven by a composition effect. Female workers are more likely to be single and, when married, are more likely to have a non-spouse proxy informant. This matters because proxy respondents are considerably less likely to report the earnings of non-spouses.

We also find that the stronger effect of proxy respondents on male workers' reported earnings reflects both compositional differences and household informational asymmetries. First, men are more likely to be located toward the top of the earnings distribution, where proxy respondents have larger negative effects on reported earnings. Second, we provide several pieces of evidence consistent with asymmetric information within households. In particular, our results reveal that men are more likely to partially conceal income from their relatives. To examine this, we analyze whether a worker's self-reported earnings vary depending on the presence of other family members during the survey. We are able to do so because of the panel structure of the ENOE, and the fact that we can approximate the presence of other household members at the time of the survey by whether they answer or not their own questionnaire. Our results show that married men self-report approximately 2.7% lower earnings when their spouse is present, while single men report similarly lower earnings when another household member is present, compared to when they are absent from the household. We find no support for an overreporting channel in which men overstate their earnings to enumerators because of social desirability.

The gender differences in the impact of proxy respondents on reported earnings could lead to the mismeasurement of common metrics. To illustrate this, we analyze the case of the gender wage gap, a widely used measure in both economics and other social sciences. To do so, we examine how the gender wage gap varies depending on whether or not we consider income reported by proxies. We find that if proxy responses are excluded, the estimated gender wage gap is 25% larger (5.1 percentage points) than the business-as-usual estimate, which relies on both self- and proxy-reported responses. In line with evidence from the

United States [Bollinger et al. \(2019\)](#), these results show that the use of proxy respondents can lead to a substantial underestimation of the gender wage gap. Our findings underscore how gender differences in income misreporting affect measures of gender inequality in developing countries.

Our paper is connected to various strands of literature. Our main contribution is to the literature on survey design [Bound et al. \(2001\)](#); [De Weerdt et al. \(2020\)](#); [Dillon et al. \(2020\)](#) and, more specifically, on the use of proxy respondents and the measurement of income.³ Despite the methodological challenges, a few papers provide evidence on the impact of proxy respondents on reported income. In the United States, [Bollinger et al. \(2019\)](#) compare earnings reported by salaried workers in the CPS to administrative tax data and, similar to us, find that proxy respondents report lower earnings for male workers (by around 6%), but not for female workers.⁴

In Paraguay and Uganda, [Masselus and Fiala \(2024\)](#) use a survey experiment to study the effect of proxy respondents on reported income using a sample of microloan applicants. They find that in Paraguay (Uganda), the income of males is 12% (between 3% and 19%) lower when it is reported by the applicant's spouse compared to when it is self-reported, although this difference is only statistically significant in Paraguay. They also find that proxy reporting leads to lower reported income for women in Uganda (no comparable data in Paraguay). However, none of these papers study the mechanisms driving the results. Hence, our main contribution to the survey literature is to provide evidence on the channels explaining the underreporting of earnings by proxy respondents.

We also contribute to the literature on gender norms and the generation and sharing of resources in the household. Existing work shows that pressure to share income with relatives or broader social networks can affect expenditure timing, and distort decisions

³ A related set of papers examine the effect of proxy respondents on different indicators. For example, several studies in developing countries find that proxy respondents have effects on the measurement of labor force participation, child labor, returns to education, assets, bargaining power, and agricultural productivity, although using ad hoc surveys with relatively small sample sizes and sparse geographical coverage [Bardasi et al. \(2011\)](#); [Dillon et al. \(2012\)](#); [Serneels et al. \(2017\)](#); [Silverio-Murillo \(2018\)](#); [Ambler et al. \(2021\)](#); [Dervisevic and Goldstein \(2023\)](#).

⁴ [Mellow and Sider \(1983\)](#) and [Bound and Krueger \(1991\)](#) compare earnings reported by salaried workers in the 1978 CPS to an employer survey and social security records, respectively and do not find evidence that proxy respondents significantly contribute to the mismeasurement of labor income. However, these studies deal with small and selected samples. Outside economics, other papers have also examined this question, finding some evidence on the relevance of proxy respondents (e.g., [Tamborini and Kim, 2013](#); [Reynolds and Wenger, 2012](#)).

about investment or labor supply [Goldberg \(2017\)](#); [Squires \(2024\)](#); [Carranza et al. \(2025\)](#). One way individuals may respond to this pressure is by hiding income from those who might otherwise claim a share of it. A series of lab-in-the-field studies, focused mainly on Africa and, to a lesser extent, South Asia, find that individuals are willing to pay to hide income from their spouses [Ashraf \(2009\)](#); [Castilla \(2019\)](#); [Pouliquen \(2023\)](#) and relatives [Jakiela and Ozier \(2016\)](#); [Boltz et al. \(2020\)](#). In a contemporaneous study, [Zhang \(2024\)](#) conducts a survey with 610 couples in Western Kenya and finds evidence of income hiding, as spouses of both sexes underestimate one another's labor income and report lower own income when their spouse is present during the interview.⁵ We contribute to this literature by providing evidence of income hiding, using a large, national survey from an upper-middle-income country. Furthermore, we link this body of literature to the literature on survey design by providing evidence on the effects of asymmetric information in the household on the measurement of income in household surveys.

Finally, our paper is connected to the literature on the gender wage gap, which is the subject of enormous attention in the social sciences [Blau and Kahn \(2017\)](#); [Goldin \(2024\)](#). We contribute to this literature by showing how the use of proxy respondents in household surveys can lead to a significant underestimation of the gender wage gap in Mexico, a large upper-middle-income country with substantial gender imbalances [Matulevich et al. \(2021\)](#).

The rest of the paper is organized as follows. Section 2 presents the data, Section 3 the estimation strategy, and Section 4 the main results and validity checks. Section 5 analyzes the mechanisms and Section 6 presents evidence on the implications for measuring the gender wage gap. Section 7 concludes.

2 | DATA AND DESCRIPTIVE STATISTICS

We use microdata from the National Survey of Occupation and Employment ENOE, which is the main source of information on the Mexican labor market. This survey is conducted by the national statistical office (INEGI) and is representative of the national population. It provides

⁵Similarly, [Chen and Collins \(2014\)](#) and [Ziparo \(2020\)](#) use ad hoc surveys in which spouses are asked about each other's incomes, and find that both men and women tend to underestimate the income reported by their spouses, consistent with income hiding.

quarterly data on occupation, net labor income, and number of hours worked per week, as well as sociodemographic information on all household members.⁶ The survey has a rotating panel design in which sampled households are followed over five consecutive quarters, maintaining 80% of the sample each quarter. While the intention is for individuals to answer their own questionnaire, if an individual is not at home, another knowledgeable household member acts as a proxy. The ENOE dataset enables the identification and characterization of the worker (individual being surveyed), the informant, and the relationship between the two. This feature is crucial for our empirical strategy because it enables us to leverage the naturally occurring variation in the survey respondent to assess differences in reported earnings by type of informant across waves.

The data span the 2005–2019 period, comprising 60 survey waves and approximately 120,000 households per quarter. We restrict the analysis to workers aged 25 to 64 at the time of the survey. We also make several sample restrictions to mitigate concerns that potential income shocks might be confounded with proxy reporting: we restrict the sample to workers who have a salaried position in every survey wave, and full-time workers who do not report working a very large number of hours, namely those whose weekly hours lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. In addition, we exclude individuals with only one interview, and those for whom there is no information on their informant, which account for a minimal fraction of cases. Approximately 32% of the interviews in this sample are self-reported, while 68% are proxy-reported, as shown in Appendix Table A.1. Finally, we restrict our sample to workers who have both proxy- and self-reported responses, accounting for slightly over 50% of the workers in this subsample. Our final sample has a total of 984,993 individual interviews conducted on 224,633 workers.⁷ We winsorize earnings at the 3rd and 97th percentiles to limit the influence of outliers.

⁶The survey is designed to measure net labor income. For salaried workers, enumerators are asked to record the disposable pay after taxes, social security contributions, and any other automatic deductions.

⁷There are 2,801,855 individuals aged 25–64 in the 60 survey waves. We exclude 1,762,960 individuals in this age bracket who are not salaried in every survey wave (i.e., they are unemployed, employers, self-employed or unpaid family workers in at least one period). We then drop 442,150 individuals whose weekly hours fall outside of the 10th to 90th percentile range, 1,021 observations where the informant is missing, and 150,357 individuals who only appear in one survey wave. Finally, we drop 221,755 individuals who have only self- or proxy-reported interviews. Our final sample is composed of 224,633 workers, for whom we have 984,993 interviews.

Table 1 reports the descriptive statistics of our main sample, separated by gender. Female workers are less likely to be head of household or be married or in an informal union, have lower monthly earnings, and work fewer hours per week compared to their male counterparts. These figures underscore the existence of gender-based disparities in both socioeconomic and labor market dimensions within our study sample. Family structures in our sample are primarily nuclear, with most households (65%) comprising couples with or without children (Appendix Table A.2). Extended families, which include additional relatives beyond the immediate family, also make up a significant portion of households (22%). The mean household size is 4.2, with minors and elderly members accounting for an average of 29% and 5% of household members, respectively.

3 | ESTIMATION STRATEGY

For a given worker, we usually observe either self- or proxy-reported earnings in national surveys. Since workers who report their own earnings and those who use a proxy respondent may differ in unobservable characteristics correlated with earnings, identifying the effect of using a proxy respondent is challenging. We overcome this difficulty by taking advantage of the panel structure of the ENOE survey. As shown in Table A.1, there is variation across survey waves in the type of respondent for 50% of salaried workers.⁸ This variation enables us to observe the labor income of the same person when they report it themselves compared to when another household member does the reporting. We estimate the following equation:

$$Y_{it} = \beta_1 Proxy\ informant_{it} + \gamma_i + \alpha_t + u_{it} \quad (1)$$

⁸Appendix Table A.3 compares the characteristics of these workers with those of workers who always self-report or always rely on proxy-reporting. Across both sexes, always self-reporting workers are older, earn more, and are substantially more likely to be head of household than always proxy-reported workers. In terms of education, always self-reporting men have higher educational attainment than always proxy-reported men, while the opposite is observed among women. Workers who have both self and proxy reports often fall between the two extremes.

Y_{it} is an outcome for worker i in period t . Our main outcomes are a dummy for whether worker i 's earnings are reported in the survey conducted in period t , the worker's monthly earnings and hourly wage (in MEX\$ of 2019, and in ln), and the weekly hours worked. We only consider workers' earnings in their main occupation. Importantly, only 5% of the workers in our sample have a second job. As the effect of proxy-reporting could vary by the gender of the worker, we estimate this equation separately for male and female workers. Our main regressor, *Proxy informant* $_{it}$, is a dummy variable that takes a value of 1 if the informant for worker i in survey period t is another household member, and 0 if the worker self-reports. Finally, γ_i and α_t are individual and survey-wave fixed effects, and u_{it} is the error term. We cluster our standard errors at the individual level.

For our coefficient of interest ($\hat{\beta}_1$) to capture the effect of proxy respondents on *reported* labor income, having a proxy respondent in a given wave should not correlate with differences in *actual* (unobserved) income. Our individual fixed effects control for any time-invariant differences in employment conditions between workers who are present at the time of the survey and workers who are not. However, in principle, there could be individual-specific determinants of labor income that vary across time and correlate with the type of respondent. This concern is partially mitigated by the fact that the workers in our sample are observed over a relatively short period (five consecutive quarters) and our sample is restricted to full-time workers who have a salaried position in every period. Nevertheless, we provide evidence in support of this assumption in Section 4.1.

It is important to point out that if proxy respondents make random errors while reporting earnings, our coefficient of interest will be zero. On the contrary, if proxies systematically misreport earnings—either by consistently over- or under-reporting them—our $\hat{\beta}_1$ will be statistically different from zero.

4 | RESULTS

Proxy respondents are less likely to report a worker's earnings than workers themselves—see columns 1 and 5 of Table 2. Among self-respondents, 90% of male and 91% of female workers report their own earnings. In contrast, proxies are 13.1 and 17.6 percentage points

less likely to report the earnings of male and female workers, respectively. The difference between these two estimates is substantial—4.5 percentage points, or roughly one-third of the estimated effect for male workers—and statistically significant at the 1% level (see Appendix Table A.4). Thus proxy reporting reduces earnings reporting more strongly for female than for male workers. The finding that proxies are less likely than workers themselves to report earnings is consistent with previous research from the United States [Bollinger et al. \(2019\)](#).

When earnings are reported (in 81% of cases), we find significant differences by respondent type in the earnings of male workers, but only minor differences for female workers. The monthly earnings reported for male workers are 7.2% lower on average when reported by a proxy, while for female workers the difference is 1.8%.⁹ The 5.4 percentage point difference in reporting gaps between male and female workers is statistically significant at the 1% level (see Appendix Table A.4). We obtain very similar results if we use nominal instead of real earnings (Appendix Table A.5), or if we use sampling weights (Appendix Table A.6).¹⁰

We find discrepancies in the reported weekly hours worked for both genders. Compared to self-reports, proxies report 0.471 more hours for male workers and 0.255 more hours for female workers; both differences are statistically significant at the 1% level. Consequently, the gap in reported hourly wages is larger than the gap in monthly earnings—8.3% vs. 7.2% for male workers, and 2.4% vs. 1.8% for female workers.

⁹To put the magnitude of our estimates in perspective, we compare the proxy-reporting gap with earnings discrepancies documented in validation studies that link workers' survey reports to employer reports or administrative records. Using matched worker and employer reports from the CPS, [Mellow and Sider \(1983\)](#) find that employer-reported wages exceed worker-reported wages by 4.8% on average. In Swedish linked survey-register data, [Kapteyn and Ypma \(2007\)](#) find a smaller average discrepancy, with survey earnings exceeding register earnings by 2.4 log points on average. Using SIPP data matched to Social Security records, [Abowd and Stinson \(2011\)](#) document larger gaps, with administrative earnings exceeding survey reports by between 10.6% and 33.3% across demographic and education groups. Our estimated 7.2% proxy-reporting gap for male workers therefore falls within the range of earnings discrepancies documented in prior validation studies.

¹⁰Non-response to the earnings questions in the ENOE has increased over time. Motivated by this pattern, we explore heterogeneity across time, by splitting the sample into three groups according to the period in which the household entered the panel (2005Q1-2009Q4, 2010Q1-2014Q4 and 2015Q1-2019Q3), and interact our coefficient of interest with dummies for the two last periods. As shown in Appendix Table A.7, the effect of proxy respondents on the likelihood of reporting earnings increases over time for both male and female workers, highlighting the role of proxy respondents in the secular increase in earnings non-response. By contrast, the effects of proxy respondents on reported monthly earnings remain stable over time, with no significant changes for male workers and only very small changes for female workers.

4.1 | Validity checks

As discussed in Section 3, the causal interpretation of our estimates requires that the type of respondent in a given survey wave is not correlated with other time-varying determinants of labor income. We perform several robustness checks to examine the validity of our identifying assumption. In particular, we show that our results are robust to limiting our sample to workers who are less likely to have experienced fluctuations in earnings throughout the survey waves. The main results from these validity checks are summarized in Figure 1.

Our full sample includes formal and informal workers, and one could be concerned that the labor income of informal workers fluctuates in ways that correlate with the worker's presence in the home. Importantly, as shown in Appendix Table A.8 we find a very similar gap in male workers' reported earnings if we limit our sample to workers who have a formal job in all periods (receive access to social security from their employers). Furthermore, our results are also very similar if we restrict the sample to workers who are only paid a salary in every period (as opposed to commissions, performance bonuses, etc.), as shown in Appendix Table A.9. We also obtain similar results if we eliminate interviews conducted in December, the month in which salaried workers receive their 13th salary (Appendix Table A.10), and if we eliminate observations in which individuals are not working their usual hours (Appendix Table A.11).

We also examine the robustness of our results to limiting our sample to workers who have been working in the same firm in all survey waves. The basic questionnaire, applied in all survey waves, does not have information on job tenure. However, interviews conducted in the first quarter of the year and in some other periods use an extended questionnaire, which asks workers about their tenure in the company in which they are currently employed.¹¹ We conduct a robustness check restricting our sample to workers who remain employed at the same firm from the quarter in which their household is initially surveyed to their final interview with the extended questionnaire. We drop interviews conducted after the last interview with the extended questionnaire, as we do not know

¹¹Besides the first quarter of every year, the extended questionnaire was also applied in all survey waves of 2005, and in the second quarters of 2006, 2007 and 2008.

whether workers changed firms after this point. Our main estimates are almost unchanged if we focus on this sample (see Table A.12 in the Appendix),¹² or if we impose all the restrictions mentioned previously (Appendix Table A.13).

A potential concern is that proxies systematically act as respondents in earlier interviews (each worker is interviewed in five survey waves), and our coefficient of interest thus captures the natural increase in earnings that occurs over time for workers who progress in their careers. Mitigating this concern, Appendix Figure A.1 shows that the estimated gap between proxy- and self-reported earnings is unaffected by the interview number in which the proxy acts as an informant. Furthermore, our results are almost unchanged if we control for survey round fixed effects (Appendix Table A.14).

Overall, our results indicate that the use of proxy respondents leads to lower reported earnings, particularly for male workers, and that these differences stem from reporting discrepancies rather than from changes in actual earnings.

4.2 | Heterogeneous proxy reporting

We now investigate the potential heterogeneity underlying our estimates of the average effect of proxy reporting on labor income. To do so, we compute the difference between the average proxy-reported and self-reported monthly earnings for each worker in our sample. Panels A and B in Figure 2 present the distribution of these reporting gaps (log differences) for male and female workers, respectively. We find substantial heterogeneity in reported earnings by informant type for both genders, with a wide dispersion in both underreporting and overreporting by proxies relative to workers' self-reports. Hence, the average effect documented above masks considerable variation in both the magnitude and the direction of the reporting gaps.

For male workers, proxies report lower earnings than workers in 56.6% of cases, higher earnings in 38.7% of cases, and identical earnings in only 4.7% of cases. For female

¹²We measure whether workers are only paid a salary, have a formal job, or remain in the same firm across all survey waves using both self- and proxy responses. Since these measures may depend on the type of informant, a potential concern is that differential reporting could bias our results. To assess this possibility, we re-estimate these regressions relying only on self-reports of these characteristics, and report these results in Panel B of Tables A.8-A.9 and A.12. Reassuringly, our main estimates remain almost unchanged compared with estimates relying on both self- and proxy-reports of the workers' job characteristics in Panel A.

workers, the corresponding shares are 49.3%, 45.6%, and 5.1%. Precision is low: the reporting gap is below 10% for only 30% of male workers and 33% of female workers, while it exceeds 50% in 15% and 12% of cases, respectively. This large variation in reporting gaps is consistent with the average effects documented in our main results, as the distribution of gaps is approximately normal but left-skewed for men, whereas for women it is centered more symmetrically around the mean.

In the previous subsection, we showed that the average effect of proxy reporting is not explained by worker characteristics associated with fluctuations in labor income. To assess whether the same holds for heterogeneous proxy reporting, we focus on a stable worker subsample—those who remain in the same firm, have a formal job, and exclusively receive a fixed salary. We find similar patterns for this group, as illustrated in Panels C and D of Figure 2. The shares of male and female workers for whom proxies report lower, higher, or exactly the same earnings as the workers themselves are very close (by up to three percentage points) to those in the full sample. Hence, as in our main results, the substantial heterogeneity in reporting gaps likely stems from reporting discrepancies rather than actual fluctuations in earnings.

Furthermore, as Figure A.2 in the Appendix shows, proxy respondents are not the only source of instability in reported earnings; self-reported labor income also varies substantially across survey waves. The average difference between the minimum and maximum self-reported earnings for the same worker is approximately 27% for both genders within the stable worker subsample. Among these stable workers, only 17% of men and 16% of women report identical earnings across survey waves. For only 36% and 34%, respectively, the gap is below 0.1 log points. Importantly, the gap between proxy- and self-reported earnings is independent of whether workers self-report their earnings consistently across survey rounds. As shown in Appendix Figure A.3, which plots the distribution of worker-level reporting differences by type of informant, the results in Figure 2 hold even when we restrict the sample to workers with little variation (less than 0.1 log points) in self-reported earnings.

Together, the findings presented here highlight an important but understudied fact: surveys may record very different earnings for the same worker performing the same job,

depending on the informant. In other words, reported earnings may not be stable even when actual earnings are.

The survey design literature has documented that income misreporting tends to be mean-reverting: low-income individuals are more likely to overstate their earnings and high-income individuals are more likely to understate them [Bound and Krueger \(1991\)](#); [Bound et al. \(1994\)](#); [Angel et al. \(2019\)](#); [Flachaire et al. \(2023\)](#). Hence, the heterogeneity in the effects of proxy reporting may depend on workers' earnings. To explore this possibility, we adapt equation (1) by fully interacting the regressor of interest with earning quintile dummies, and using the real monthly earnings (in ln) as the dependent variable. To avoid categorizing workers incorrectly, we use self-reported earnings to compute earning quintiles. Figure 3 presents our results.

We find that the effect of proxy respondents varies substantially across the distribution of earnings. Proxy-reported earnings are higher than self-reported earnings in the lowest quintile (17.4% for males, and 11.6% for females). On the contrary, proxy respondents have a negative effect on reported earnings in the top three quintiles, with larger effects toward the top of the distribution. For male workers, proxy-reported earnings are 5.8% lower than self-reported earnings in the third quintile, 15.0% in the fourth quintile, and 27.7% in the top quintile. For female workers, proxy-reported earnings are 4.2% lower than self-reported earnings in the third quintile, 9.6% in the fourth quintile, and 24.8% in the top quintile. Overall, these results show that proxy reporting introduces mean-reverting measurement error in earnings.

These findings also indicate that part of the reason we observe a larger gap between the average self- and proxy-reported earnings of male workers is the overrepresentation of males in higher earnings levels. However, it is worth noting that gender gaps in proxy reporting persist even among workers within the same earnings group. This remains the case even if we split the sample into deciles (Appendix Figure A.4) or ventiles (Appendix Figure A.5). This suggests that, beyond earnings, the gap between self- and proxy-reported earnings for male workers is shaped by gender-specific household dynamics.

5 | WHY IS THERE A GAP BETWEEN SELF- AND PROXY-REPORTING?

We now turn to the question of why the effect of proxy reporting on both the probability of reporting earnings and reported monthly earnings differs by gender. Motivated by the household economics literature, we consider information asymmetries within households as a potentially relevant mechanism [Deschênes et al. \(2020\)](#). However, an alternative explanation is that workers tend to overstate their earnings when reporting directly to enumerators. In this section, we examine which of these mechanisms is at play.

5.1 | Asymmetric information and income hiding

To study the asymmetric information mechanism, we first examine whether the effect of proxy respondents varies by how well the proxy knows other characteristics about the worker's job. Specifically, we distinguish between proxies who are better informed—accurately reporting the worker's occupation and industry, which occurs in 70% of cases—and those who misreport at least one of these categories (as verified against the worker's self-reported data). To ensure accurate measurement of proxy knowledge, we restrict our sample to workers who have remained employed at the same firm across all survey waves, thereby holding constant their occupation and industry.¹³ We interact equation (1) with a dummy variable for whether the proxy respondent is not knowledgeable. Table 3 presents results.

Uninformed proxies are slightly less likely than informed proxies to report earnings for both male and female workers (see columns 1 and 5). For male—but not female—workers, uninformed proxies report lower earnings on average than informed proxies (columns 2 and 4). However, the magnitude of the reporting gap between the two types of proxies is small compared to the gap between informed proxies and self-reports (0.8% vs. 7.2% for males). In other words, even household members who are relatively well informed about the worker's job systematically underreport the worker's earnings.

Because information asymmetries may depend on the relationship between the

¹³We also drop 13% of workers who either do not self-report their occupation and industry, or whose responses are inconsistent across survey waves.

worker and the informant, we examine whether the proxy reporting gap varies by the worker-proxy relationship, distinguishing between single and married workers.¹⁴ Same-sex couples account for 0.12% of the latter group. Among married workers, spouses act as proxies in 79% of cases, followed by children (17%) and other household members (4%) (see Panel B in Appendix Table A.15). Males are more likely than females to be married (83.6% vs. 56.3%), and among those who are married, they are more likely to have their spouse as a proxy informant (84% vs. 67%). For single workers, children are the most common proxies (48%), followed by parents, siblings, and others (19%, 17%, and 18% respectively).

Panel A of Table 4 presents results of our estimates of the effect of proxy reporting for married workers and Panel B for single workers. For married workers, spouses are more likely than other household members to report earnings. In particular, relative to self-reports, men's spouses are 7.8 percentage points less likely to report earnings, while women's spouses are only 0.7 percentage points less likely to do so. By contrast, other household members are 25.9 and 23.1 percentage points less likely to report the earnings of married men and married women, respectively. Among single workers, proxy respondents are 23.2 percentage points less likely than workers to report men's earnings and 21.8 percentage points less likely to report women's earnings.¹⁵

These results help explain why the average effect of proxy reporting on the probability of reporting earnings is larger for female than for male workers: as noted above, female workers are more likely to be single and, when married, are more likely to have a non-spouse proxy informant.

Regarding reported earnings, the gap between self- and proxy-reporting among married men is larger when the spouse is the informant (7.3%) than when it is another household member (5%). For married women, the corresponding gaps are negligible: 0.7% for spouse informants and 0.03% for other household members, with the latter not statistically significant. Among single workers, proxy-reported earnings are lower than self-reported earnings both for men (8.3%) and women (3.1%), although the gap is smaller

¹⁴The sample of married workers is composed of individuals who are married or in an informal union in every period and for whom we can identify their spouse in the data. For ease of exposition, we use the term married to refer to these individuals. We identify cohabiting couples using data on the relationship with the household head for individuals who report being married or in a union. We could not unambiguously identify the spouse in 7% of the cases where the worker is married or in an informal union.

¹⁵In Appendix Table A.16, we show the results of regressions allowing the effects to vary by proxy relationship.

for women. Overall, the use of proxy informants leads to lower reported earnings for men, both married and single, and single women, although the effect is smaller for the latter group.

We then test whether workers intentionally conceal their earnings from family members by analyzing if for a given worker, self-reported earnings vary with the presence of other household members at the time of the survey. For single workers, we add to our estimating equation a dummy variable for whether the worker self-reports his/her earnings and at least one other household member is present at the time of the survey; the comparison group is workers who self-report when nobody else is around. We consider other household members to be absent if a proxy gives the interview for them. Accordingly, we have an upper bound of actual presence in the interview, as others might be present in the dwelling, but not necessarily in the room where the interview takes place. For married workers, we distinguish between the presence of their spouse and other household members. We thus add to our estimating equation a dummy variable for whether the worker self-reports and his/her spouse is present at the time of the survey, and a dummy for whether the worker self-reports and another household member (but not his/her spouse) is present at the time of the survey. For identification, we assume that for workers who self-report, the presence of other household members during the survey is uncorrelated with time-varying determinants of the workers' labor income. The results of these estimations are presented in Table 5.

We find that both male and female workers are less likely to report their earnings when their spouse or other household members are present in the dwelling at the time of the interview, as shown in columns 1 and 5. This pattern is consistent with income hiding within the household. Among married workers, men are 3.0 percentage points less likely to report their earnings when their spouse is present and 6.2 percentage points less likely when other household members are present. For women, the corresponding figures are 2.4 and 6.6 percentage points, respectively. Among single workers, men are 5.9 percentage points and women 5.0 percentage points less likely to report their earnings when other household members are present.

When we examine reported earnings, we find that married male workers report 2.7%

lower earnings when their spouse is present than when they self-report with no one else present. This difference is statistically significant at the 1% level.¹⁶ When other household members are present, there is no difference in the self-reported earnings of married men. In the case of single male workers, their self-reported earnings are 2.4% lower when others are present compared to when the worker answers the questionnaire with nobody else around. More than half of the single male workers in our sample live with their parents (Appendix Table A.2) and could possibly have incentive to hide their income to retain discretion over their spending. For female workers, we find no difference in self-reported earnings between interviews conducted when the worker is alone and those conducted when another household member is present.

Hence, for men, the evidence on reported earnings is consistent with married workers hiding their earnings from their spouses and single workers hiding their earnings from other household members. However, the earnings reported by proxies are lower than those self-reported by workers in the presence of others. This might be due to our measure of the presence of others, which probably overestimates actual presence. Furthermore, proxy informants may also underreport the earnings of workers if they systematically underestimate them, regardless of whether or not the worker has the intention of hiding his/her income.

Differences in bargaining power may lead to higher income hiding and asymmetric information within couples Doss (2013). We examine whether the difference between the self- and proxy-reported earnings of married workers varies with bargaining power in the couple, which we proxy using differences in educational attainment.¹⁷ For the sample of married workers, we fully interact our estimating equation, distinguishing between spouse and other informants, with dummies for whether the respondent has less, equal,

¹⁶ Masselus and Fiala (2024) find that the reported earnings of male workers are 11–14% higher when their spouse is not present (although these differences are not statistically significant); this could explain why the gap between self- and partner-reported earnings for male workers is higher in their study.

¹⁷ Educational attainment is measured in the household questionnaire, which is typically completed by a single household member on behalf of all household members during the first interview. In our sample of married workers, this variable is self-reported in 37% of cases, reported by spouses in 45%, by other household members in 10%, and in the remaining 8% we cannot identify who reported this specific variable. This may introduce measurement error. However, educational attainment is mainly reported by the worker and his/her spouse, and spouses are plausibly better informed about their partner's education than other household members. Moreover, our analysis relies only on relative differences in educational attainment between spouses, rather than exact years of schooling, which likely makes this measure less susceptible to reporting error. We therefore believe that any resulting measurement error is unlikely to materially affect our results.

or more education than his/her spouse. Panel A of Figure 4 reports the coefficients of the interaction effects for the dummy of the spouse informant. For men, workers self-report higher earnings than their spouses report for them across all relative-education groups. The size of this proxy-reporting gap varies with the worker's education relative to his spouse's, and is largest when the worker has more education than his spouse. For women, we do not find a proxy-reporting gap among workers with less than or the same level of education as their spouses. However, we do find a significant proxy-reporting effect when the worker has more education than her spouse—although the effect is substantially smaller than for men. This evidence is consistent with greater income hiding in couples characterized by asymmetric bargaining power.

Asymmetric information within households, and particularly within couples, is closely linked to gender norms: more egalitarian norms are associated with lower informational asymmetries (Deschênes et al., 2020). However, the ENOE does not collect information on gender norms, which limits our ability to assess whether they are the mechanism underlying income hiding in our context.

5.2 | Do workers overstate their earnings to enumerators?

As discussed in Section 4.2, the survey design literature shows that low-income individuals tend to overstate their wages and high-income individuals to understate them. If lower-income workers overreport their income with respect to proxies and higher-income workers underreport it, we would expect to observe negative proxy reporting gaps for low-income workers and positive proxy reporting gaps for high-income workers. Instead, we find the opposite pattern.

A related hypothesis is that men overstate their earnings to conform to societal expectations of being the primary breadwinner Bertrand et al. (2015); Slotwinski and Roth (2020). If this were the case, the gap in reported earnings would be larger for men whose spouse is employed. We explore this hypothesis by testing for heterogeneous effects across the employment status of spouses in the subsample of married workers. We find no evidence in support of this hypothesis, as our estimations show that the difference between the self-

reported earnings of male workers and their earnings reported by their spouse is *smaller* when the worker's spouse is employed (Panel B of Figure 4). Furthermore, we find that female workers self-report higher earnings than their proxies report for them when their spouse is unemployed or out of the labor force. This pattern is consistent with income hiding among those women who may have stronger incentives or greater capacity to conceal income from their partners.¹⁸

We further explore the hypothesis that men overstate their earnings to conform to societal expectations by studying whether self-reported earnings vary with the interviewer's gender. Specifically, we include in the model a dummy variable equal to 1 if the interviewer is female (0 if male) and an interaction term between this dummy and the indicator for proxy respondent. If male workers overstate their earnings when interviewed by women (or by men), we would expect the coefficient on the female-interviewer dummy to be significantly different from zero. We estimate the model using data from 2006 to 2019—we exclude observations from 2005 because interviewer gender is not recorded, and drop 0.01% of cases from 2006 onward where this information is missing. Overall, 43% of interviews in the resulting sample were conducted by women—see Appendix Table A.17. Table 6 reports results.¹⁹

We find no evidence that men's self-reported earnings vary with the interviewer's gender. The coefficient on the female-interviewer dummy is close to zero (0.002). Proxy respondents for male workers tend to report slightly lower earnings (the estimated coefficient is -0.010) when interviewed by women. For female workers, neither self-reports nor proxy reports differ statistically significantly by interviewer gender.

In sum, our results indicate that the observed discrepancies in workers' reported

¹⁸One alternative explanation for the smaller average effect of proxy reporting on women's reported earnings is that some women may conceal part of their income both from other household members and from interviewers. If the degree of underreporting were similar in both cases, proxy reporting would not necessarily generate a detectable difference in reported earnings. Although we cannot test this mechanism directly because we do not observe actual earnings, several patterns described in this section are difficult to reconcile with it. In particular, the effect of proxy reporting on women's reported earnings is negative when women appear to have greater bargaining power within the household—namely, when they have higher relative education than their spouses—and when they may have stronger incentives to conceal income from other household members, such as when their own earnings are higher or when their spouse is unemployed. These patterns suggest that, when incentives to hide income are stronger, women may conceal more from relatives than from interviewers.

¹⁹We had access to information on the interviewer's gender—which is not included in the publicly available datasets—through INEGI's Microdata Lab. Interested researchers can apply for access by following the institutional process described at <https://www.inegi.org.mx/programas/enoe/15ymas/>.

earnings by respondent type arise from proxy respondents misreporting the workers' earnings, rather than workers misreporting their own earnings. The gender gap in the reported earnings of workers by respondent type is explained by two effects. First, there is a composition effect. Males tend to have higher earnings than females, and higher earnings are associated with negative gaps between proxy and self-reported earnings. Second, there is an asymmetry of information effect, partly explained by males having a higher propensity to hide a portion of their income from other household members, who serve as proxy informants.

6 | IMPLICATIONS FOR MEASURING GENDER INEQUALITY

The results presented in Section 4 show that the reported earnings of male workers are substantially lower when reported by a proxy, while the gap between the self- and proxy-reported earnings of female workers is small. Therefore, the widespread use of proxy respondents could lead to the mismeasurement of common metrics derived from labor force surveys such as ENOE. In this section, we focus on the effect of proxy-reporting on the gender wage gap.²⁰

We estimate the gender wage gap using the following standard equation:

$$\ln(\text{Hourly Wage}_{it}) = \beta_0 + \beta_1 \text{Female}_i + X_{it}\rho + \gamma_t + U_{it} \quad (2)$$

Hourly Wage_{it} is the hourly wage of worker *i* in quarter *t* and *Female_i* is a dummy for whether worker *i* is female (0 if male). *X_{it}* is a vector of covariates that controls for age, educational attainment, urban residence, and state of residence, and γ_t captures survey-wave fixed effects. The parameter of interest β_1 captures the average difference between wages earned by female and male workers conditional on the characteristics included in

²⁰Proxy reporting may affect income-based statistics such as the gender wage gap both by changing the probability that earnings are reported and by changing reported earnings conditional on reporting. Our analysis in this section focuses on the latter margin.

X_{it} .²¹ Additionally, to examine how the gender wage gap varies depending on whether or not the earnings reported by proxies are considered, we interact Female_i with the indicator for whether the observation is proxy-reported, and include this indicator by itself as well.

We conduct these estimations using the same sample as in our previous estimations. Because this sample differs from other workers in the economy (e.g., salaried workers who always self- or proxy-report their earnings, part-time workers, and the self-employed), the estimates reported here should not be interpreted as estimates of the gender wage gap in the overall workforce. Rather, they are intended to illustrate the extent to which proxy respondents may distort the measurement of this indicator.

Table 7 presents our results. The estimated gender wage gap when using both self- and proxy-reporting is -12.8% (column 1), whereas using only self-reported income leads to an estimated gender wage gap of -16.0% (Female coefficient in column 2 in the same table). Restricting the analysis to proxy-reported observations yields a smaller gender wage gap in absolute value: -10.9% , compared with -16.0% when using only self-reported earnings. The difference of 5.1 percentage points between these two estimates is statistically significant at the 1% level. Therefore, if proxy respondents are excluded, the estimated gender wage gap is 25% larger than the business-as-usual estimate, which uses both self- and proxy-reported responses.²² These results indicate that the use of proxy respondents can lead to a substantial underestimation of the gender wage gap.

As previously mentioned, we conducted these estimations on the sample of workers with both proxy- and self-reported earnings. To assess the generality of these results, we now perform an additional exercise on a more representative sample. We include workers who always self-report and workers who always rely on proxy respondents, and compare the estimated gender wage gap with and without imputing proxy-reported hourly wages. For workers for whom we observe both self- and proxy-reported wages (52% of the sample), we replace proxy-reported hourly wages with the worker's average self-reported hourly

²¹One can estimate the gender wage gap with fewer or more controls depending on the purpose of the exercise. A more parsimonious specification controlling only for survey-wave and age controls will simply document overall differences in wages by gender. Including additional controls such as occupation and industry may be useful if the goal is to compare men and women with more similar job characteristics. Our baseline specification takes an intermediate approach. Reassuringly, the main conclusion is unchanged under either alternative, as shown in Appendix Table A.18.

²²We obtain similar results if we separate by region, as shown in Appendix Figure A.6.

wage. For workers observed only through proxy reports, we impute hourly wages using the average wage of observationally similar workers in the same quarter, matched on gender, state, urban/rural status, educational attainment and age group, and if possible, by industry, and occupation.²³ As shown in Appendix Table A.19, this exercise yields a substantially larger estimated gender wage gap than the one based on the uncorrected data: the estimated gap increases by about 5 percentage points from 11.6% to 16.5%. This pattern is consistent with our main results and further suggests that relying on proxy-reported wages leads to a substantial understatement of gender differences in earnings. When respondent identity is available, imputing proxy-reported wages using workers' own self-reports and information from observationally similar workers may be a promising strategy to attenuate the resulting measurement error.

7 | CONCLUSIONS

Our study brings new evidence on the impact of proxy respondents on reported earnings, and particularly on the underlying mechanisms. Using the panel structure of the Mexican labor force survey, we demonstrate substantial differences between self-reporting and proxy-reporting of earnings, with proxy respondents reporting lower earnings for male workers, with smaller effects for female workers. Our analysis of mechanisms reveals that this gap is due to asymmetric information within households and a tendency for men to hide their income from other household members. Furthermore, we show that the gender differences between self- and proxy-reporting of labor income can lead to a significant underestimation of the gender wage gap.

The use of proxy respondents helps to lower costs, but, as we show, it can introduce measurement error in key variables such as labor income, because social norms affect the flow of information within the household. Researchers can benefit from assessing how the use of proxy respondents affects their results, particularly if gender plays a role in their

²³We managed to match exactly on quarter, gender, state, urban/rural status, age group, educational attainment, occupation, and industry in 45% of cases. When no exact match was available, we broadened the match by first dropping industry (30%), then matching on industry but not occupation (7%), then dropping both occupation and industry (11%). We drop 6% of cases where we could not find a match using these criteria.

analysis. To this end, it is necessary for statistical offices to include in their survey datasets information about the identity of the respondents. For this article, we assessed 11 labor force surveys (or equivalents) carried out by national statistical offices from Latin American countries. Although all of these surveys use proxies, it was only possible to know if a proxy gave the interview in six out of 11 of the surveys (in Colombia, Ecuador, Mexico, Paraguay, Peru, and Uruguay, but not in Argentina, Bolivia, Brazil, Chile, and Costa Rica). Overall, our research underscores the importance of understanding the role of proxy respondents in survey data collection to ensure the accuracy and reliability of socioeconomic indicators.

ACKNOWLEDGEMENTS

We are grateful for comments from Karen Macours, Martin Rossi, Pauline Rossi, Adan Silverio Murillo, Joe Vecchi, and participants at seminars at CAF, CIDE, Colegio de México, DIAL, Paris School of Economics, Universidad Nacional de la Plata, and Universidad Torcuato Di Tella, and at the ECINEQ-LAC Meeting, the Gender and Development Workshop at Nanterre University, and the Max Weber Conference at the EUI. We thank INEGI's Microdata Lab and the ENOE team for access to non-public data on interviewer gender.

| Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT for writing assistance, language editing, and coding support. The authors reviewed and edited the output as needed and take full responsibility for the content of the published article.

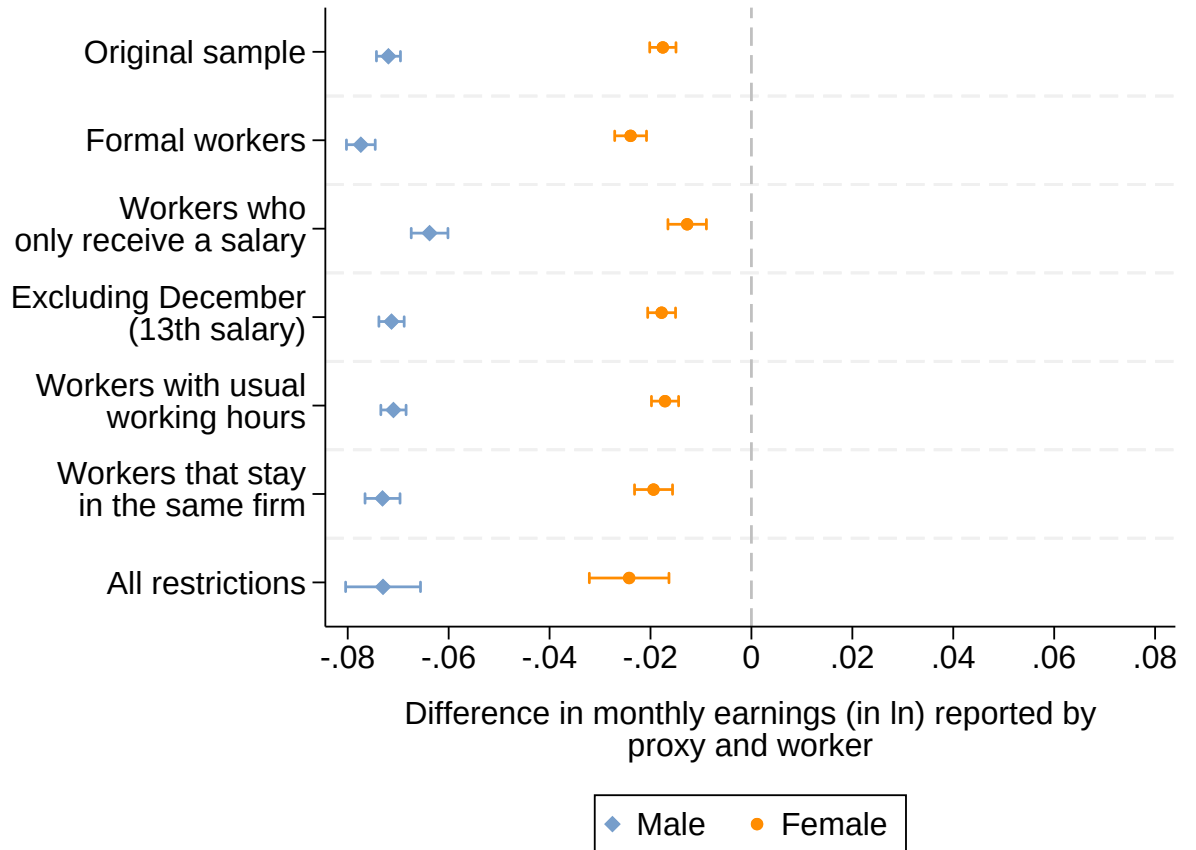
REFERENCES

- Abowd, J. M. and Stinson, M. H. (2011) Estimating measurement error in SIPP annual job earnings: A comparison of census bureau survey and SSA administrative data. *Working Paper SEHSD-WP2011-19 / SIPP-WP-252*, U.S. Census Bureau, Washington, DC.
- (2013) Estimating measurement error in annual job earnings: a comparison of survey and administrative data. *Review of Economics and Statistics*, **95**, 1451–1467.
- Ambler, K., Doss, C., Kieran, C. and Passarelli, S. (2021) He says, she says: Spousal disagreement in survey measures of bargaining power. *Economic Development and Cultural Change*, **69**, 765–788.
- Angel, S., Disslbacher, F., Humer, S. and Schnetzer, M. (2019) What did you really earn last year?: explaining measurement error in survey income data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **182**, 1411–1437.
- Ashraf, N. (2009) Spousal control and intra-household decision making: An experimental study in the philippines. *American Economic Review*, **99**, 1245–1277.
- Bardasi, E., Beegle, K., Dillon, A. and Serneels, P. (2011) Do labor statistics depend on how and to whom the questions are asked? results from a survey experiment in tanzania. *The World Bank Economic Review*, **25**, 418–447.
- Bertrand, M., Kamenica, E. and Pan, J. (2015) Gender identity and relative income within households. *Quarterly Journal of Economics*, **130**, 571–614.
- Blau, F. D. and Kahn, L. M. (2017) The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, **55**, 789–865.
- Bollinger, C. R., Hirsch, B. T., Hokayem, C. M. and Ziliak, J. P. (2019) Trouble in the tails? what we know about earnings nonresponse 30 years after lillard, smith, and welch. *Journal of Political Economy*, **127**, 2143–2185.
- Boltz, M., Marazyán, K. and Villar, P. (2020) Is informal redistribution costly? evidence from a lab-in-the-field experiment in senegal. *The World Bank Economic Review*, **34**, S72–S78.
- Bound, J., Brown, C., Duncan, G. J. and Rodgers, W. L. (1994) Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, **12**, 345–368.
- Bound, J., Brown, C. and Mathiowetz, N. (2001) Measurement error in survey data. In *Handbook of econometrics*, vol. 5, 3705–3843. Elsevier.

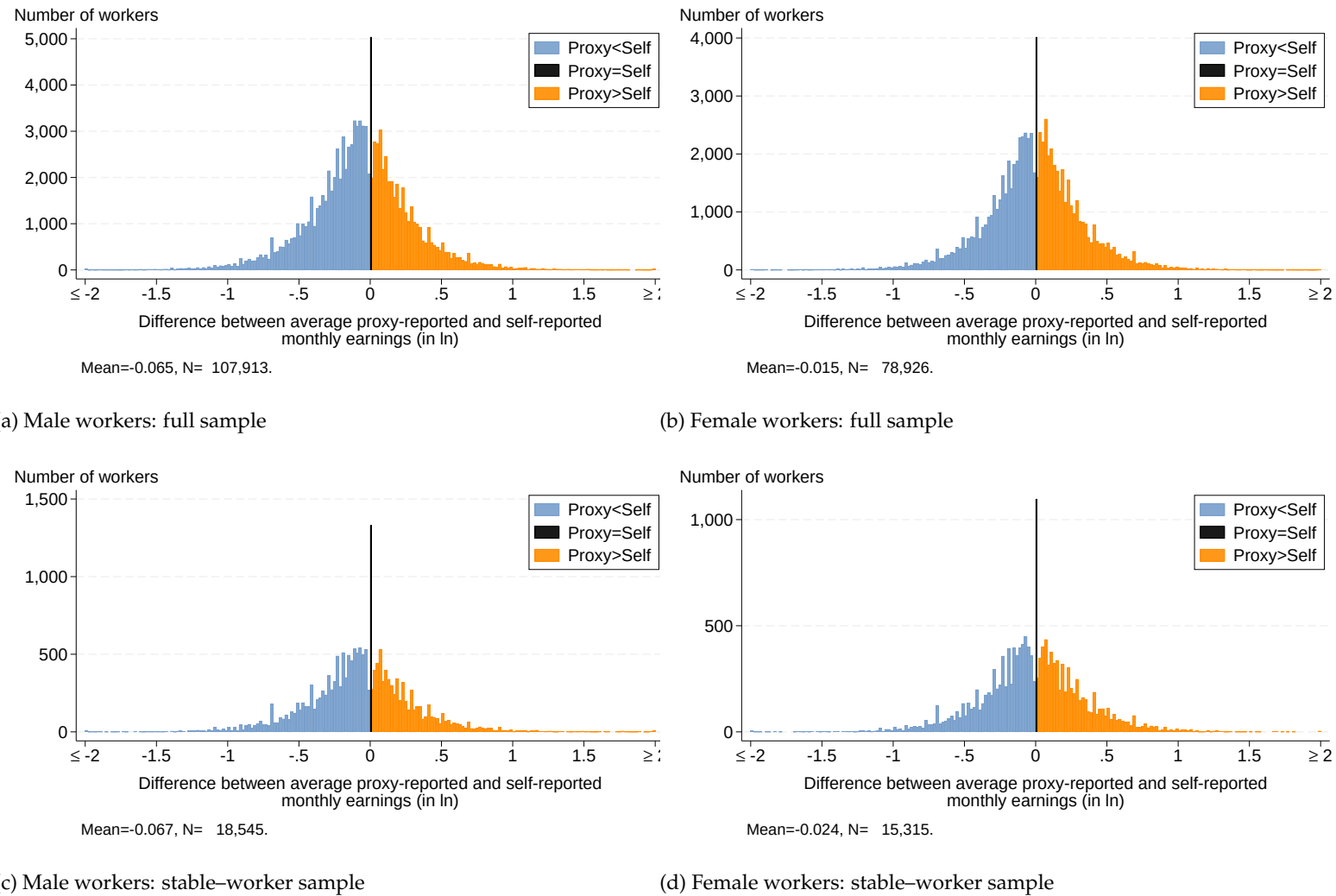
- Bound, J. and Krueger, A. B. (1991) The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, **9**, 1–24.
- Carranza, E., Donald, A., Grosset-Touba, F. and Kaur, S. (2025) The Social Tax: Redistributive Pressure and Labor Supply. *Econometrica*, **93**, 2273–2308.
- Castilla, C. (2019) What’s yours is mine, and what’s mine is mine: Field experiment on income concealing between spouses in india. *Journal of Development Economics*, **137**, 125–140.
- Chen, J. J. and Collins, L. A. (2014) Let’s talk about the money: Spousal communication, expenditures, and farm production. *American Journal of Agricultural Economics*, **96**, 1272–1290.
- De Weerd, J., Gibson, J. and Beegle, K. (2020) What can we learn from experimenting with survey methods? *Annual Review of Resource Economics*, **12**, 431–447.
- Dervisevic, E. and Goldstein, M. (2023) He said, she said: The impact of gender and marriage perceptions on self and proxy reporting of labor. *Journal of Development Economics*, **161**, 103028.
- Deschênes, S., Dumas, C. and Lambert, S. (2020) Household resources and individual strategies. *World Development*, **135**, 105075.
- Dillon, A., Bardasi, E., Beegle, K. and Serneels, P. (2012) Explaining variation in child labor statistics. *Journal of Development Economics*, **98**, 136–147.
- Dillon, A., Karlan, D., Udry, C. and Zinman, J. (2020) Good identification, meet good data. *World Development*, **127**, 104796.
- Doss, C. (2013) Intrahousehold bargaining and resource allocation in developing countries. *The World Bank Research Observer*, **28**, 52–78.
- Flachaire, E., Lustig, N. and Vigorito, A. (2023) Underreporting of top incomes and inequality: A comparison of correction methods using simulations and linked survey and tax data. *Review of Income and Wealth*, **69**, 1033–1059.
- Goldberg, J. (2017) The Effect of Social Pressure on Expenditures in Malawi. *Journal of Economic Behavior & Organization*, **143**, 173–185.
- Goldin, C. (2024) Nobel Prize Lecture. URL: <https://www.nobelprize.org/prizes/economic-sciences/2023/goldin/lecture/>.
- Jakiela, P. and Ozier, O. (2016) Does africa need a rotten kin theorem? experimental evidence from village economies. *The Review of Economic Studies*, **83**, 231–268.

- Kapteyn, A. and Ypma, J. Y. (2007) Measurement error and misclassification: a comparison of survey and administrative data. *Journal of Labor Economics*, **25**, 513–551.
- Masselus, L. and Fiala, N. (2024) Whom to ask? testing respondent effects in household surveys. *Journal of Development Economics*, 103265.
- Matulevich, R., Galeano, E. C. and Servian, D. M. (2021) Mexico - country gender scorecard. Country Gender Scorecards FY22 Washington, D.C.: World Bank Group, available at <http://documents.worldbank.org/curated/en/159831645690741028/Mexico-Country-Gender-Scorecard>.
- Mellow, W. and Sider, H. (1983) Accuracy of response in labor market surveys: Evidence and implications. *Journal of Labor Economics*, **1**, 331–344.
- Meyer, B. D., Mok, W. K. C. and Sullivan, J. X. (2015) Household surveys in crisis. *Journal of Economic Perspectives*, **29**, 199–226.
- Pouliquen, V. (2023) This business is mine! individual property rights and intra-household relationships. Working paper.
- Reynolds, J. and Wenger, J. B. (2012) He said, she said: the gender wage gap according to self and proxy reports in the current population survey. *Social Science Research*, **41**, 392–411.
- Serneels, P., Beegle, K. and Dillon, A. (2017) Do returns to education depend on how and whom you ask? *Economics of Education Review*, **60**, 5–19.
- Silverio-Murillo, A. (2018) Wife vs. husband: Does it matter who answers the survey? Working Paper.
- Slotwinski, M. and Roth, A. (2020) Gender norms and income misreporting within households. *ZEW-Centre for European Economic Research Discussion Paper*.
- Squires, M. (2024) Kinship Taxation as an Impediment to Growth: Experimental Evidence from Kenyan Microenterprises. *The Economic Journal*, **134**, 2558–2579.
- Tamborini, C. R. and Kim, C. (2013) Are proxy interviews associated with biased earnings reports? marital status and gender effects of proxy. *Social Science Research*, **42**, 499–512.
- U.S. Census Bureau (2019) Current population survey design and methodology. Technical Paper 77.
- Zhang, S. (2024) Hidden in Plain Sight: Asymmetric Information and Hidden Income within the Household. Working paper.
- Ziparo, R. (2020) Why do spouses communicate: love or interest? a model and some evidence from cameroon. Working paper.

FIGURES AND TABLES

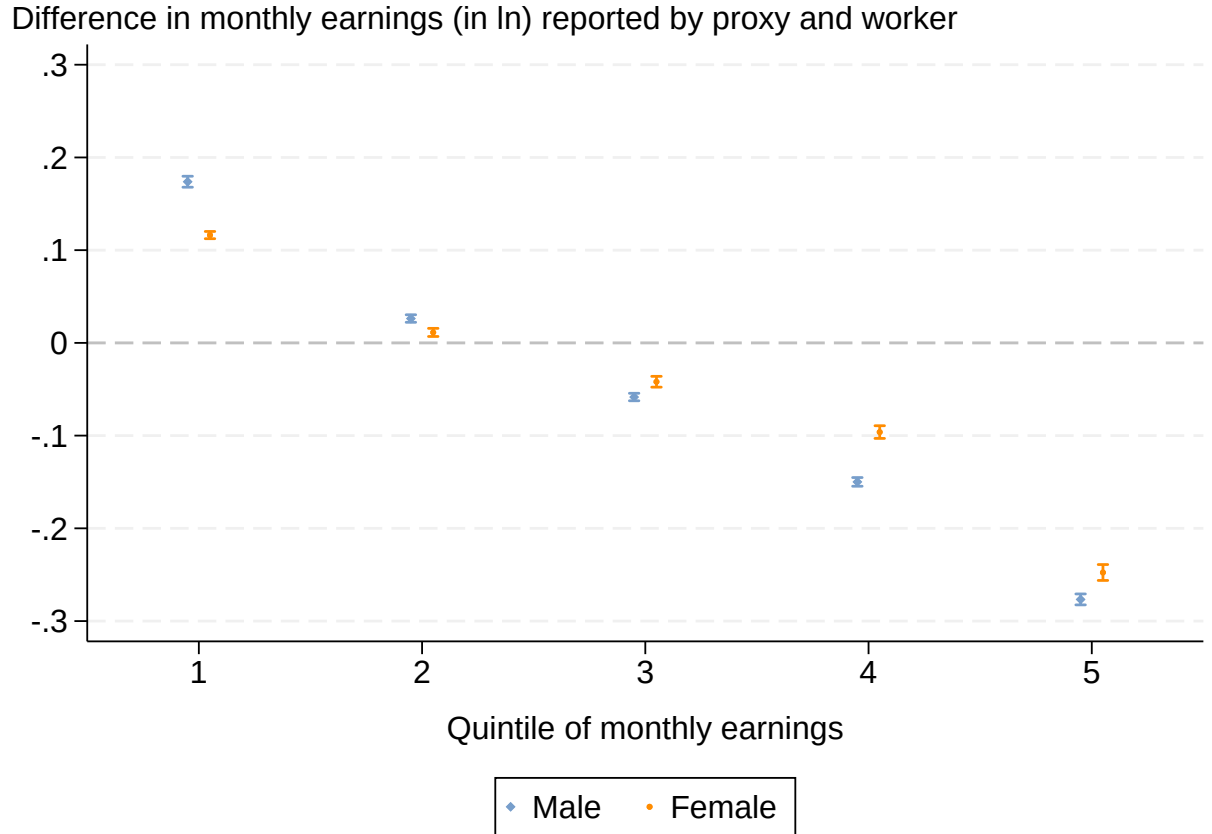
FIGURE 1 Differences in reported earnings by type of respondent – summary of validity checks

Note: This figure presents the main coefficients and 95% confidence intervals for the estimation of equation (1) for different subsamples. The dependent variable is the monthly earnings (in MEX\$ of 2019 and in ln). The estimates of regressions in which the worker is male (female) are presented in blue (orange). The original sample includes all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported; we exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The sample of formal workers is restricted to those that receive access to social security from their employers in every period. The sample of salaried workers who only receive a salary is limited to workers who have a salaried position for which they only receive a salary in all periods. The sample excluding December excludes observations from interviews conducted in December. The sample of workers with usual working hours is restricted to observations for which the reported hours correspond to the number of hours the worker habitually works. The sample of workers who stay in the same firm is limited to workers who stay at the same firm throughout survey waves. The “All restrictions” sample applies all of the preceding restrictions simultaneously.

FIGURE 2 Distribution of worker-level differences between proxy-reported and self-reported earnings

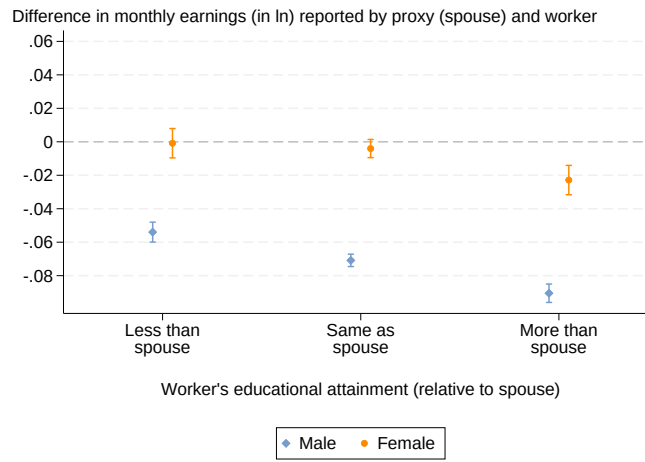
Note: The full sample is composed of all salaried workers aged 25–64 who have both self- and proxy-reported earnings. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The stable worker subsample further restricts to workers who remain in the same firm, have a formal job in every period, and exclusively receive a fixed salary. This figure presents the histogram of worker-level log differences between average proxy-reported and average self-reported monthly earnings.

FIGURE 3 Differences in reported earnings by type of respondent – heterogeneous effects by workers' earnings

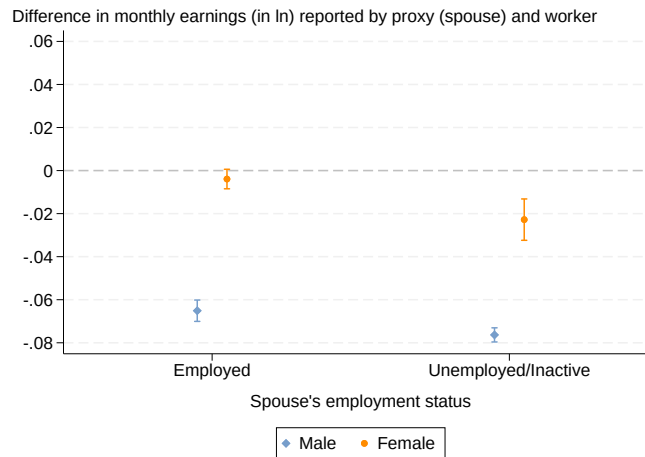


Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. This figure presents the main coefficients and 95% confidence intervals for the estimation of equation (1) fully interacted with earning quintile dummies, constructed using self-reported earnings. The dependent variable is the monthly earnings (in MEX\$ of 2019 and in ln). The estimates of regressions in which the worker is male (female) are presented in blue (orange).

FIGURE 4 Differences in reported earnings by type of respondent – heterogeneous effects by worker's and spouse's relative education and spouse's employment status



(a) Heterogeneous effects by worker's and spouse's relative education



(b) Heterogeneous effects by spouse's employment status

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. Finally, the sample is restricted to workers who are married or in a union and cohabit with their spouse in all periods. We exclude partnered workers for whom we cannot identify the spouse. Panel A presents the main coefficients and 95% confidence intervals for the estimation of equation (1), distinguishing by whether the proxy is the respondent's spouse or another household member, and fully interacted with dummies for whether the worker has less, the same, or more education than his/her partner. Panel B presents the analogous estimation, in which the interaction is with dummies for whether the worker's partner is employed or unemployed/inactive. We only report the coefficients of the interaction effects for the dummy of the partner informant. The dependent variable is monthly earnings (in MEX\$ of 2019 and in ln). The estimates of regressions in which the worker is male (female) are presented in blue (orange).

TABLE 1 Descriptive statistics

	Male				Female			
	Mean	SD	Min	Max	Mean	SD	Min	Max
<i>Demographics</i>								
Age	39.503	9.541	25.000	64.000	39.755	8.959	25.000	64.000
Head of household	0.794	0.404	0.000	1.000	0.263	0.440	0.000	1.000
Married or informal union	0.836	0.370	0.000	1.000	0.563	0.496	0.000	1.000
<i>Educational attainment</i>								
Incomplete primary	0.060	0.237	0.000	1.000	0.048	0.213	0.000	1.000
Complete primary	0.138	0.345	0.000	1.000	0.127	0.333	0.000	1.000
Complete lower secondary	0.353	0.478	0.000	1.000	0.372	0.483	0.000	1.000
Complete upper secondary or more	0.449	0.497	0.000	1.000	0.453	0.498	0.000	1.000
<i>Labor market outcomes</i>								
Hours worked per week	45.975	6.529	30.000	60.000	43.781	6.556	30.000	60.000
Monthly earnings (in MEX\$ of 2019)	9206.059	5514.197	2461.522	31648.146	7887.255	4957.462	2461.522	31648.146
Hourly wage (in MEX\$ of 2019)	51.547	33.425	10.256	263.735	46.810	31.977	10.256	263.735
<i>Job characteristics</i>								
Formal employment	0.833	0.373	0.000	1.000	0.825	0.380	0.000	1.000
Primary sector	0.040	0.195	0.000	1.000	0.007	0.085	0.000	1.000
Secondary sector	0.384	0.486	0.000	1.000	0.210	0.407	0.000	1.000
Tertiary sector	0.575	0.494	0.000	1.000	0.782	0.413	0.000	1.000
Observations	561,300				423,693			
Individuals	128,415				96,218			

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. We split the sample by gender.

TABLE 2 Differences in reported earnings by type of respondent

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.131*** (0.001)	-0.072*** (0.001)	-0.083*** (0.001)	0.471*** (0.018)	-0.176*** (0.001)	-0.018*** (0.001)	-0.024*** (0.001)	0.255*** (0.019)
Observations	561,300	460,329	460,329	561,300	423,693	342,378	342,378	423,693
Individuals	128,415	125,412	125,412	128,415	96,218	94,208	94,208	96,218
R ²	0.462	0.788	0.790	0.524	0.453	0.816	0.825	0.595
Dep. var. mean (self)	0.899	9.043	3.847	45.559	0.911	8.843	3.695	43.484
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable taking a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 3 Differences in reported earnings by type of respondent – heterogeneous effects by proxy knowledge

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.122*** (0.002)	-0.072*** (0.002)	-0.081*** (0.002)	0.442*** (0.031)	-0.163*** (0.002)	-0.019*** (0.002)	-0.025*** (0.002)	0.241*** (0.031)
Proxy respondent $\times$ Uninformed proxy	-0.015*** (0.003)	-0.008** (0.003)	-0.009*** (0.003)	-0.005 (0.047)	-0.024*** (0.004)	-0.000 (0.004)	0.001 (0.005)	-0.077 (0.057)
Observations	237,836	197,102	197,102	237,836	183,781	150,240	150,240	183,781
Individuals	67,718	65,735	65,735	67,718	52,212	50,827	50,827	52,212
R ²	0.493	0.811	0.813	0.571	0.483	0.833	0.842	0.636
Dep. var. mean (self)	0.909	9.045	3.852	45.434	0.920	8.842	3.699	43.265
P-value (Effect for uninformed proxy=0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. We also restrict our sample to individuals who work for the same company in the period between their first interview and their last interview with the extended questionnaire, and workers who self-report the same occupation and industry in all survey waves. We drop interviews conducted after the last interview with the extended questionnaire. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressors are a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports, as well as the interaction of this regressor with a dummy for whether the proxy respondent does not correctly report the workers' occupation or industry. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 4 Differences in reported earnings by type of respondent – by marital status

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
<i>Panel A: Married sample</i>								
Spouse informant	-0.078*** (0.001)	-0.073*** (0.001)	-0.084*** (0.001)	0.496*** (0.021)	-0.091*** (0.002)	-0.007*** (0.002)	-0.011*** (0.002)	0.149*** (0.030)
Other informant	-0.259*** (0.003)	-0.050*** (0.003)	-0.058*** (0.003)	0.339*** (0.040)	-0.231*** (0.003)	-0.003 (0.003)	-0.010*** (0.003)	0.288*** (0.042)
Observations	436,724	365,805	365,805	436,724	208,279	174,430	174,430	208,279
Individuals	98,581	96,442	96,442	98,581	47,121	46,317	46,317	47,121
R ²	0.471	0.790	0.793	0.529	0.449	0.821	0.829	0.618
Dep. var. mean (self)	0.900	9.063	3.868	45.496	0.912	8.887	3.751	42.968
P-value (spouse=other)	0.000	0.000	0.000	0.000	0.000	0.203	0.805	0.003
<i>Panel B: Single sample</i>								
Proxy respondent	-0.232*** (0.003)	-0.083*** (0.003)	-0.095*** (0.003)	0.530*** (0.046)	-0.218*** (0.002)	-0.031*** (0.002)	-0.039*** (0.002)	0.306*** (0.031)
Observations	92,053	68,514	68,514	92,053	184,948	143,649	143,649	184,948
Individuals	21,713	21,015	21,015	21,713	41,990	40,918	40,918	41,990
R ²	0.470	0.775	0.777	0.502	0.463	0.806	0.813	0.566
Dep. var. mean (self)	0.891	8.977	3.781	45.566	0.907	8.796	3.634	44.096
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. In Panel A, the sample is restricted to workers who are married or in a union and cohabit with their spouse in all periods. The married sample excludes partnered workers for whom we cannot identify the spouse. In Panel B, the sample is restricted to workers that are single in all periods. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressors in Panel A are dummy variables for whether the informant is the worker's spouse or another household member (the omitted category is self-reporting by the worker). The main regressor in Panel B is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 5 Differences in reported earnings by type of respondent – by marital status and presence of others when self-reporting

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
<i>Panel A: Married sample</i>								
Self-reporting and spouse is present	-0.030*** (0.002)	-0.027*** (0.003)	-0.029*** (0.003)	0.128*** (0.045)	-0.024*** (0.004)	0.001 (0.004)	-0.002 (0.004)	0.123** (0.062)
Self-reporting and others are present	-0.062*** (0.007)	0.002 (0.007)	-0.001 (0.008)	0.086 (0.115)	-0.066*** (0.007)	-0.004 (0.007)	-0.005 (0.007)	0.018 (0.110)
Spouse informant	-0.085*** (0.001)	-0.078*** (0.002)	-0.089*** (0.002)	0.519*** (0.023)	-0.096*** (0.002)	-0.007*** (0.002)	-0.012*** (0.002)	0.165*** (0.031)
Other informant	-0.266*** (0.003)	-0.054*** (0.003)	-0.062*** (0.003)	0.362*** (0.041)	-0.237*** (0.003)	-0.003 (0.003)	-0.011*** (0.003)	0.301*** (0.043)
Observations	436,724	365,805	365,805	436,724	208,279	174,430	174,430	208,279
Individuals	98,581	96,442	96,442	98,581	47,121	46,317	46,317	47,121
R ²	0.471	0.790	0.793	0.529	0.449	0.821	0.829	0.618
Dep. var. mean (self alone)	0.896	9.079	3.886	45.426	0.911	8.892	3.757	42.904
P-value (spouse proxy=spouse present)	0.000	0.000	0.000	0.000	0.000	0.026	0.016	0.501
P-value (other proxy=others present)	0.000	0.000	0.000	0.018	0.000	0.926	0.435	0.010
<i>Panel B: Single sample</i>								
Self-reporting and others are present	-0.059*** (0.006)	-0.024*** (0.006)	-0.024*** (0.006)	0.036 (0.092)	-0.050*** (0.004)	-0.004 (0.004)	-0.005 (0.004)	0.067 (0.065)
Proxy respondent	-0.245*** (0.004)	-0.089*** (0.004)	-0.101*** (0.004)	0.538*** (0.051)	-0.228*** (0.002)	-0.032*** (0.002)	-0.040*** (0.002)	0.318*** (0.033)
Observations	92,053	68,514	68,514	92,053	184,948	143,649	143,649	184,948
Individuals	21,713	21,015	21,015	21,713	41,990	40,918	40,918	41,990
R ²	0.471	0.775	0.777	0.502	0.464	0.806	0.813	0.566
Dep. var. mean (self alone)	0.893	8.995	3.800	45.537	0.908	8.802	3.642	44.027
P-value (proxy=others present)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. In Panel A, the sample is restricted to workers who are married or in a union and cohabit with their spouse in all periods. The married sample excludes partnered workers for whom we cannot identify the spouse. In Panel B, the sample is restricted to workers that are single in all periods. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 6 Differences in reported earnings by type of respondent – by interviewer's gender

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.123*** (0.001)	-0.068*** (0.002)	-0.078*** (0.002)	0.469*** (0.025)	-0.162*** (0.002)	-0.016*** (0.002)	-0.022*** (0.002)	0.252*** (0.026)
Female interviewer	-0.012*** (0.002)	0.002 (0.002)	0.000 (0.002)	0.129*** (0.031)	-0.009*** (0.002)	-0.002 (0.002)	-0.003 (0.002)	0.085*** (0.030)
Proxy × Female interviewer	-0.028*** (0.002)	-0.010*** (0.003)	-0.011*** (0.003)	0.001 (0.038)	-0.047*** (0.003)	-0.003 (0.003)	-0.003 (0.003)	0.005 (0.040)
Observations	527,062	428,761	428,761	527,062	399,624	320,402	320,402	399,624
Individuals	122,767	119,563	119,563	122,767	92,311	90,117	90,117	92,311
R ²	0.464	0.788	0.790	0.525	0.457	0.816	0.825	0.596
Dep. var. mean (self)	0.895	9.040	3.844	45.572	0.907	8.840	3.691	43.521
P-value (Female interviewer + Proxy × Female interviewer)	0.000	0.000	0.000	0.000	0.000	0.024	0.006	0.002
P-value (Proxy + Proxy × Female interviewer)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We include observations from the 1st quarter of 2006 to the 4th quarter of 2019. We exclude workers who do not have a salaried position in all periods and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable taking a value of 1 if the informant is another household member, and 0 if the worker self-reports. We also include as regressors a dummy variable equal to 1 if the interviewer is female (0 if male), as well as an interaction term between this dummy and the proxy-respondent dummy. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 7 Gender wage gap

	(1)	(2)
Female	-0.128*** (0.002)	-0.160*** (0.002)
Female $\times$ Proxy respondent		0.051*** (0.002)
Observations	802,319	802,319
R ²	0.286	0.291
Controls	✓	✓
Time FE	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The dependent variable in all regressions is the worker's hourly wage (in MEX\$ of 2019 and in ln). Our main regressor is a dummy for whether the worker is female. We also control for age, age squared, educational attainment dummies, a dummy for whether the worker lives in an urban area, state fixed effects, and survey-wave fixed effects. In column 2, we also include a regressor for whether the observation is proxy reported, and the interaction between the female dummy and this variable. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

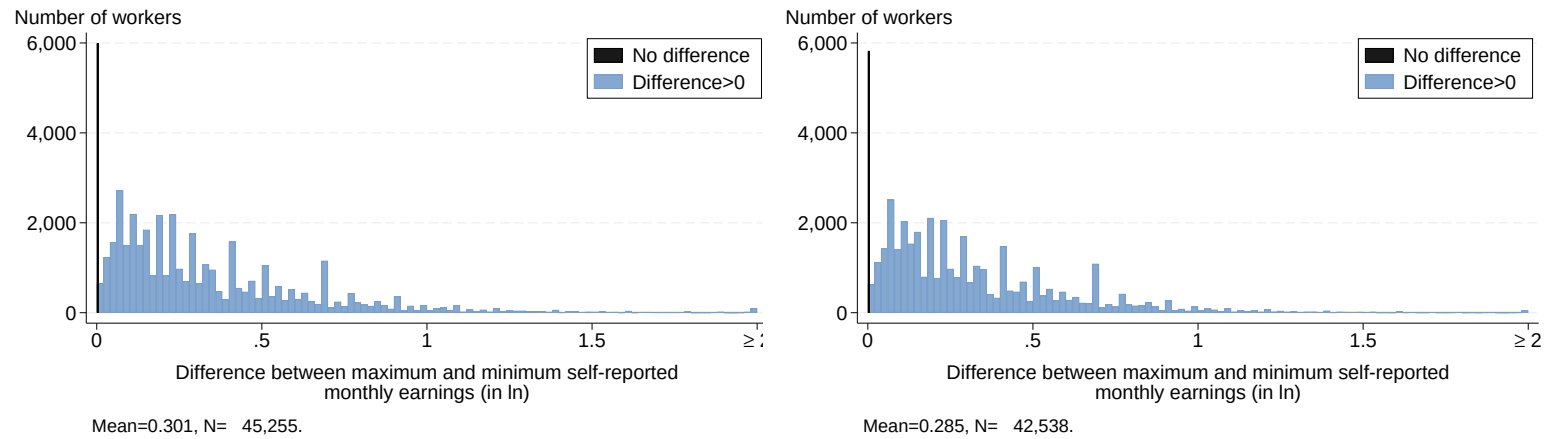
ONLINE APPENDIX

Appendix A | APPENDIX FIGURES AND TABLES

FIGURE A.1 Differences in reported earnings by type of respondent – heterogeneous effects by interview number

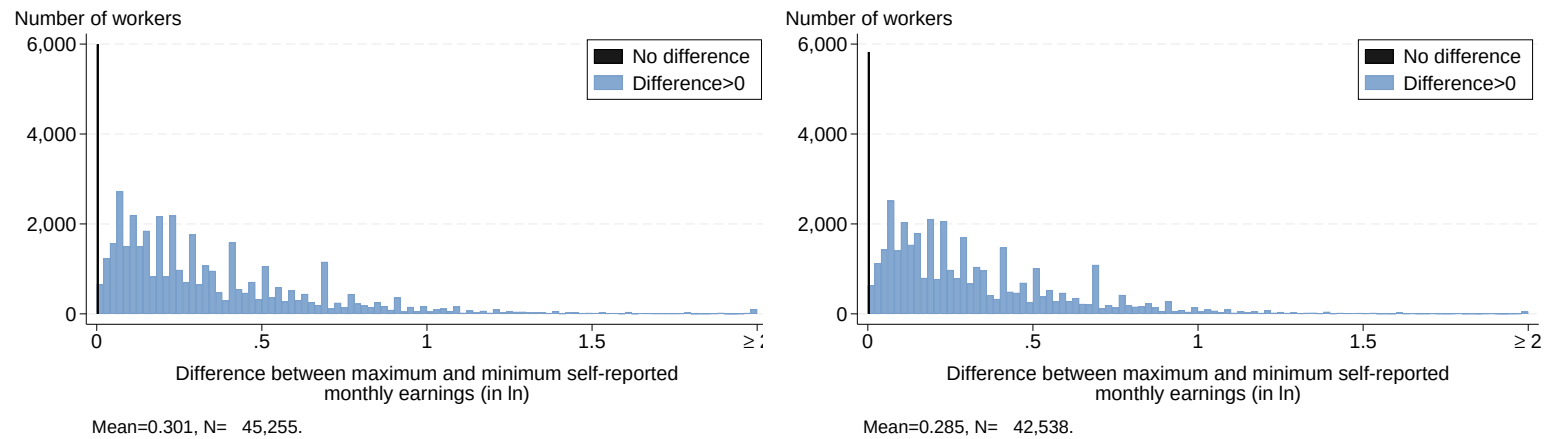


Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. This figure presents the main coefficients and 95% confidence intervals for the estimation of equation (1) fully interacted with dummies for the interview number from which the observation comes. The dependent variable is the monthly earnings (in MEX\$ of 2019 and in ln). The estimates of regressions in which the worker is male (female) are presented in blue (orange).

FIGURE A.2 Distribution of worker-level differences between maximum and minimum self-reported earnings

(a) Male workers: full sample

(b) Female workers: full sample

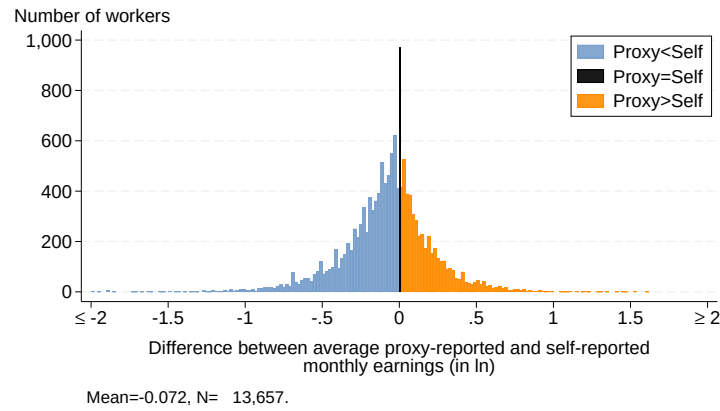


(c) Male workers: stable-worker sample

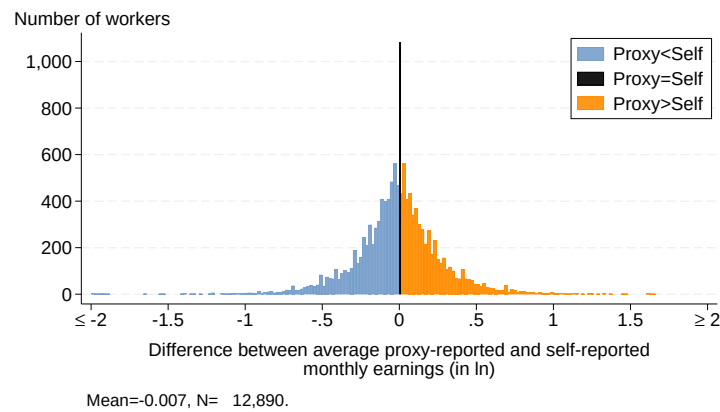
(d) Female workers: stable-worker sample

Note: The sample is composed of salaried workers aged 25–64 who have both self- and proxy reported earnings and have at least two observations with self-reported earnings. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The stable worker subsample further restricts to workers who remain in the same firm, have a formal job in every period, and exclusively receive a fixed salary. This figure presents the histogram of worker-level log differences between maximum and minimum self-reported monthly earnings.

FIGURE A.3 Distribution of worker-level differences between proxy-reported and self-reported earnings – sample with workers with stable self-reported earnings



(a) Male workers



(b) Female workers

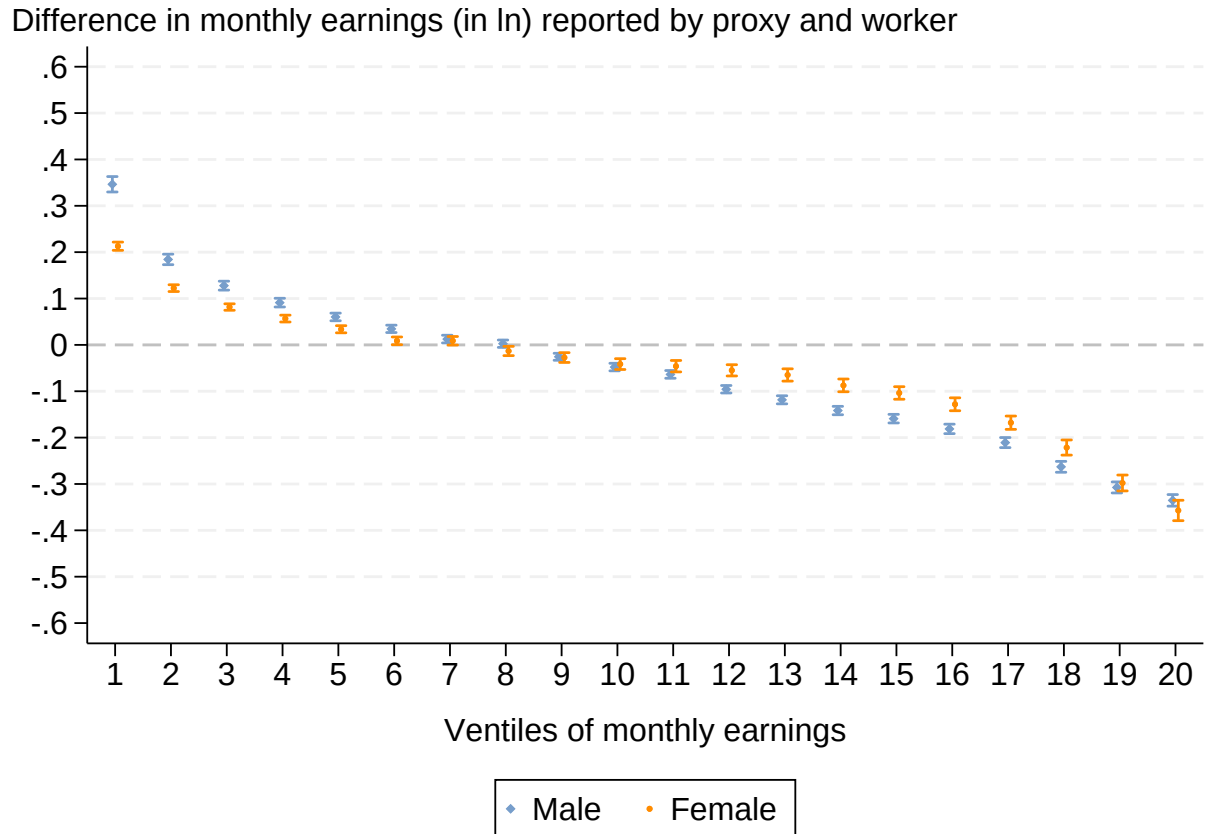
Note: The sample is composed of all salaried workers aged 25–64 who have both self- and proxy-reported earnings, have at least two observations with self-reported earnings, and whose log differences between maximum and minimum self-reported monthly earnings is smaller or equal than 0.1 log points. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. This figure presents the histogram of worker-level log differences between average proxy-reported and average self-reported monthly earnings.

FIGURE A.4 Differences in reported earnings by type of respondent – heterogeneous effects by workers' earnings (in deciles)

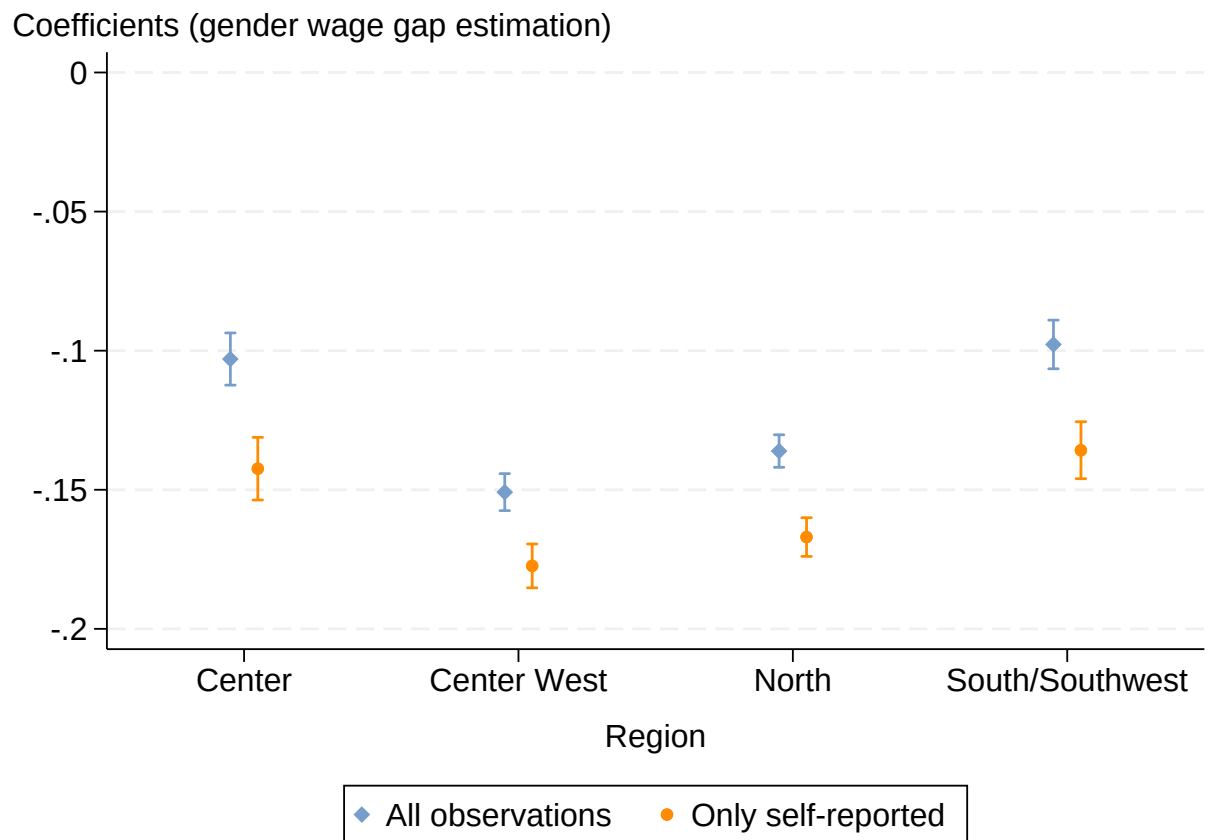


Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. This figure presents the main coefficients and 95% confidence intervals for the estimation of equation (1) fully interacted with earnings decile dummies, constructed using self-reported earnings. The dependent variable is the monthly earnings (in MEX\$ of 2019 and in ln). The estimates of regressions in which the worker is male (female) are presented in blue (orange).

FIGURE A.5 Differences in reported earnings by type of respondent – heterogeneous effects by workers' earnings (in ventiles)



Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. This figure presents the main coefficients and 95% confidence intervals for the estimation of equation (1) fully interacted with earnings ventiles dummies, constructed using self-reported earnings. The dependent variable is the monthly earnings (in MEX\$ of 2019 and in ln). The estimates of regressions in which the worker is male (female) are presented in blue (orange).

FIGURE A.6 Gender wage gap – heterogeneity by region

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. This figure presents the results of a regression where the dependent variable is the worker's hourly wage (in MEX\$ of 2019 and in ln), and the main regressor is a dummy for whether the worker is female, interacted with regional dummies. We also control for age, age squared, educational attainment dummies, a dummy for whether the worker lives in an urban area, state fixed effects, and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. We report the estimated coefficient for the Female dummy in each region using the entire sample in blue, and using only self-reported observations in orange, with the corresponding 95% confidence interval.

TABLE A.1 Type of informant

	All	Male	Female
<i>Panel A: Observations</i>			
Self-reporting	0.321	0.269	0.400
Proxy respondent	0.679	0.731	0.600
Observations	1,836,567	1,099,393	737,174
<i>Panel B: Individuals</i>			
Always self-reporting	0.115	0.086	0.160
Always proxy-reporting	0.381	0.436	0.299
Both self- and proxy-reporting	0.503	0.479	0.540
Observations	446,388	268,266	178,122

Note: The sample is composed of all salaried workers aged 25–64. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave, and workers who only appear in one survey wave. Panel A shows the share of observations by informant type, and Panel B shows the distribution of the type of informant.

TABLE A.2 Household characteristics of workers

	All			Male			Female		
	All	Married	Single	All	Married	Single	All	Married	Single
<i>Demographics</i>									
Household size	4.191	4.169	4.054	4.181	4.188	3.807	4.204	4.131	4.177
Share of individuals aged 17 or less	0.288	0.335	0.184	0.299	0.346	0.098	0.273	0.311	0.227
Share of individuals aged between 18 and 64	0.667	0.649	0.713	0.665	0.640	0.781	0.669	0.668	0.679
Share of individuals aged 65 or more	0.046	0.016	0.103	0.036	0.014	0.121	0.058	0.021	0.094
Average number of children per women	1.936	1.808	2.204	2.021	1.845	2.956	1.828	1.730	1.895
Male head of households	0.769	0.942	0.419	0.892	0.953	0.680	0.606	0.918	0.289
Share of total time spent on chores done by women	0.838	0.813	0.927	0.781	0.788	0.560	0.894	0.852	0.963
<i>Head of household educational attainment</i>									
Incomplete primary	0.111	0.064	0.199	0.100	0.066	0.214	0.124	0.058	0.192
Complete primary	0.174	0.142	0.235	0.166	0.144	0.236	0.186	0.137	0.234
Complete lower secondary	0.327	0.349	0.283	0.337	0.360	0.253	0.313	0.328	0.298
Complete upper secondary or more	0.388	0.445	0.282	0.397	0.430	0.296	0.376	0.478	0.275
<i>Labor market outcomes</i>									
Per capita household labor income (PPP dollars)	305.979	316.227	289.897	293.741	285.586	337.118	322.191	380.475	266.394
Labor income of working-age men (PPP dollars)	548.645	606.414	392.669	592.191	618.632	492.723	477.850	580.535	312.331
Labor income of working-age women (PPP dollars)	329.625	332.972	328.431	232.287	240.100	197.473	450.834	527.335	373.913
Share of working-age men who work	0.862	0.890	0.777	0.914	0.914	0.916	0.776	0.841	0.665
Share of working-age women who work	0.632	0.600	0.711	0.472	0.469	0.485	0.831	0.873	0.789
<i>Household type</i>									
Single person	0.005	0.000	0.013	0.005	0.000	0.023	0.004	0.000	0.008
Non-family	0.015	0.000	0.042	0.014	0.000	0.059	0.017	0.000	0.033
Family	0.978	1.000	0.941	0.979	1.000	0.912	0.975	1.000	0.956
Nuclear	0.756	0.879	0.566	0.795	0.886	0.565	0.704	0.865	0.567
Couple without children	0.269	0.399	0.009	0.335	0.421	0.014	0.181	0.353	0.007
Couple with children	0.376	0.480	0.195	0.410	0.465	0.267	0.332	0.512	0.159
Male single parent	0.014	0.000	0.044	0.018	0.000	0.091	0.009	0.000	0.020
Female single parent	0.097	0.000	0.318	0.033	0.000	0.193	0.182	0.000	0.381
Extended	0.222	0.121	0.375	0.184	0.114	0.347	0.271	0.134	0.389
Nuclear with at least another relative	0.181	0.120	0.248	0.152	0.114	0.185	0.219	0.134	0.279
Head of household with at least another relative	0.041	0.001	0.127	0.032	0.001	0.162	0.052	0.001	0.110
Unknown	0.022	0.000	0.059	0.021	0.000	0.088	0.025	0.000	0.044
Observations	984,993	645,003	277,001	561,300	436,724	92,053	423,693	208,279	184,948
Individuals	224,633	145,702	63,703	128,415	98,581	21,713	96,218	47,121	41,990

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. We split the sample by gender and by marital status of the worker (columns). The married sample consists of workers who are married or in a union and cohabit with their spouse in all periods, and the single sample is restricted to workers that are single in all periods. The married sample excludes partnered workers for whom we cannot identify the spouse. All variables (rows) describe characteristics of the households in which workers reside. Where variables are further disaggregated by sex — such as labor income or labor force participation — they capture all household members of the relevant sex, including the focal worker when their sex matches the category being measured. "Non-family" refers to households in which no members are related to the head.

TABLE A.3 Descriptive statistics by type of informant

	Male			Female		
	Always self-reporting	Always proxy	Self and proxy	Always self-reporting	Always proxy	Self and proxy
<i>Demographics</i>						
Age	40.120	38.562	39.503	39.744	36.678	39.755
Head of household	0.939	0.600	0.794	0.647	0.098	0.263
Married or informal union	0.473	0.750	0.836	0.396	0.348	0.563
<i>Educational attainment</i>						
Incomplete primary	0.062	0.077	0.060	0.058	0.042	0.048
Complete primary	0.115	0.169	0.138	0.127	0.117	0.127
Complete lower secondary	0.277	0.345	0.353	0.345	0.321	0.372
Complete upper secondary or more	0.545	0.408	0.449	0.470	0.518	0.453
<i>Labor market outcomes</i>						
Hours worked per week	45.551	46.937	45.975	43.532	44.831	43.781
Monthly earnings (in MEX\$ of 2019)	10302.805	8313.703	9206.059	8274.591	7432.582	7887.255
Hourly wage (in MEX\$ of 2019)	58.313	45.353	51.547	49.367	42.809	46.810
<i>Job characteristics</i>						
Formal employment	0.814	0.758	0.833	0.788	0.788	0.825
Primary sector	0.037	0.048	0.040	0.010	0.007	0.007
Secondary sector	0.313	0.414	0.384	0.188	0.222	0.210
Tertiary sector	0.648	0.536	0.575	0.801	0.770	0.782
Observations	78,059	460,034	561,300	104,699	208,782	423,693
Individuals	22,938	116,913	128,415	28,572	53,332	96,218

Note: The sample is composed of all salaried workers aged 25–64. We exclude workers who do not have a salaried position in all periods, those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave, and workers who only appear in one survey wave. We split the sample by gender and by the types of informant the worker has across survey waves.

TABLE A.4 Differences in reported earnings by type of respondent – Interaction with worker's gender

	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.176*** (0.001)	-0.017*** (0.001)	-0.024*** (0.001)	0.256*** (0.019)
Proxy respondent $\times$ Male worker	0.045*** (0.002)	-0.054*** (0.002)	-0.059*** (0.002)	0.214*** (0.026)
Observations	984,993	802,707	802,707	984,993
Individuals	224,633	219,620	219,620	224,633
R ²	0.458	0.805	0.807	0.567
Dep. var. mean (self)	0.905	8.949	3.776	44.591
P-value (Proxy + Interaction)	0.000	0.000	0.000	0.000
Individual FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The dependent variable in column 1 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 3 is the real monthly (hourly) earnings in ln. The dependent variable in column 4 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports, as well as the interaction with a dummy for whether the worker is male. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.5 Differences in reported earnings by type of respondent – nominal earnings

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.131*** (0.001)	-0.072*** (0.001)	-0.083*** (0.001)	0.471*** (0.018)	-0.176*** (0.001)	-0.018*** (0.001)	-0.024*** (0.001)	0.255*** (0.019)
Observations	561,300	460,329	460,329	561,300	423,693	342,378	342,378	423,693
Individuals	128,415	125,412	125,412	128,415	96,218	94,208	94,208	96,218
R ²	0.462	0.799	0.799	0.524	0.453	0.826	0.832	0.595
Dep. var. mean (self)	0.899	8.764	3.568	45.559	0.911	8.566	3.418	43.484
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the nominal monthly (hourly) earnings in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.6 Differences in reported earnings by type of respondent – using sampling weights

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.135*** (0.002)	-0.070*** (0.002)	-0.081*** (0.002)	0.469*** (0.031)	-0.184*** (0.002)	-0.014*** (0.002)	-0.020*** (0.002)	0.176*** (0.034)
Observations	561,300	460,329	460,329	561,300	423,693	342,378	342,378	423,693
Individuals	128,415	125,412	125,412	128,415	96,218	94,208	94,208	96,218
R ²	0.480	0.797	0.795	0.499	0.468	0.822	0.827	0.570
Dep. var. mean (self)	0.874	9.010	3.811	45.658	0.881	8.820	3.666	43.709
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the real monthly (hourly) earnings in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects, and use the sampling weights provided by ENOE. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.7 Differences in reported earnings by type of respondent – Heterogeneous effects by survey year

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.102*** (0.002)	-0.072*** (0.002)	-0.083*** (0.002)	0.476*** (0.031)	-0.133*** (0.002)	-0.021*** (0.002)	-0.026*** (0.002)	0.208*** (0.033)
Proxy respondent × Period 2010:Q1 to 2014:Q4	-0.034*** (0.003)	0.003 (0.003)	0.004 (0.003)	-0.031 (0.044)	-0.054*** (0.003)	0.002 (0.003)	0.000 (0.003)	0.066 (0.047)
Proxy respondent × Period 2015:Q1 to 2019:Q3	-0.056*** (0.003)	-0.002 (0.003)	-0.004 (0.003)	0.017 (0.044)	-0.075*** (0.003)	0.009*** (0.003)	0.007* (0.003)	0.077 (0.047)
Observations	561,300	460,329	460,329	561,300	423,693	342,378	342,378	423,693
Individuals	128,415	125,412	125,412	128,415	96,218	94,208	94,208	96,218
R ²	0.462	0.788	0.790	0.524	0.454	0.816	0.825	0.595
Dep. var. mean (self)	0.899	9.043	3.847	45.559	0.911	8.843	3.695	43.484
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The dependent variable in column 1 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 3 is the real monthly (hourly) earnings in ln. The dependent variable in column 4 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports, as well as interactions with dummies for the period in which the worker was interviewed. We split the sample into three groups according to the period in which the household entered the panel (2005:Q1-2009:Q4, 2010:Q1-2014:Q4 and 2015:Q1-2019:Q3), interact with dummies for the two last period (the omitted category is the 2004Q2-2009Q2 period). We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.8 Differences in reported earnings by type of respondent – formal workers

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
<i>Panel A: Formal according to worker or proxy</i>								
Proxy respondent	-0.136*** (0.001)	-0.077*** (0.001)	-0.086*** (0.002)	0.368*** (0.020)	-0.180*** (0.002)	-0.024*** (0.002)	-0.029*** (0.002)	0.195*** (0.021)
Observations	406,491	328,885	328,885	406,491	306,777	245,010	245,010	306,777
Individuals	92,363	89,986	89,986	92,363	69,452	67,863	67,863	69,452
R ²	0.461	0.783	0.788	0.551	0.452	0.808	0.823	0.627
Dep. var. mean (self)	0.891	9.119	3.929	45.274	0.903	8.938	3.793	43.323
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
<i>Panel B: Formal according to worker</i>								
Proxy respondent	-0.138*** (0.001)	-0.081*** (0.001)	-0.090*** (0.001)	0.412*** (0.019)	-0.183*** (0.002)	-0.027*** (0.002)	-0.032*** (0.002)	0.196*** (0.020)
Observations	459,741	371,111	371,111	459,741	337,556	268,158	268,158	337,556
Individuals	104,379	101,673	101,673	104,379	76,348	74,554	74,554	76,348
R ²	0.461	0.780	0.785	0.541	0.453	0.807	0.822	0.618
Dep. var. mean (self)	0.892	9.106	3.914	45.370	0.903	8.926	3.779	43.410
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all formal (receive access to social security from their employers in every period) salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. In Panel A, we identify formal using proxy- and self-reported responses to this question. In Panel B, we identify formal workers using only self-reported responses to this question. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.9 Differences in reported earnings by type of respondent – workers who are only paid a salary

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
<i>Panel A: Only paid a salary according to worker or proxy</i>								
Proxy respondent	-0.135*** (0.002)	-0.064*** (0.002)	-0.075*** (0.002)	0.501*** (0.029)	-0.182*** (0.002)	-0.013*** (0.002)	-0.020*** (0.002)	0.258*** (0.029)
Observations	233,023	188,065	188,065	233,023	202,063	160,091	160,091	202,063
Individuals	54,719	53,159	53,159	54,719	46,830	45,619	45,619	46,830
R ²	0.476	0.814	0.818	0.553	0.468	0.834	0.841	0.602
Dep. var. mean (self)	0.890	9.034	3.855	44.872	0.899	8.841	3.715	42.617
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
<i>Panel B: Only paid a salary according to worker</i>								
Proxy respondent	-0.131*** (0.002)	-0.060*** (0.002)	-0.071*** (0.002)	0.497*** (0.024)	-0.177*** (0.002)	-0.010*** (0.002)	-0.017*** (0.002)	0.283*** (0.025)
Observations	333,824	269,126	269,126	333,824	265,769	210,155	210,155	265,769
Individuals	77,090	74,991	74,991	77,090	60,963	59,443	59,443	60,963
R ²	0.470	0.806	0.809	0.540	0.461	0.828	0.836	0.597
Dep. var. mean (self)	0.889	9.033	3.848	45.090	0.899	8.841	3.710	42.841
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who are paid only a salary in every period (instead of commissions, performance bonuses, etc.) and who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. In Panel A, we identify workers who are only paid a salary using proxy- and self-reported responses to this question. In Panel B, we identify workers who are only paid a salary using only self-reported responses to this question. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.10 Differences in reported earnings by type of respondent – interviews conducted in January–November

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.128*** (0.001)	-0.071*** (0.001)	-0.082*** (0.001)	0.472*** (0.019)	-0.173*** (0.001)	-0.018*** (0.001)	-0.024*** (0.001)	0.252*** (0.021)
Observations	521,489	429,589	429,589	521,489	390,456	316,722	316,722	390,456
Individuals	128,412	124,630	124,630	128,412	96,214	93,588	93,588	96,214
R ²	0.476	0.795	0.797	0.535	0.468	0.822	0.830	0.606
Dep. var. mean (self)	0.902	9.037	3.841	45.588	0.912	8.838	3.689	43.526
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. We also exclude interviews conducted in December. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.11 Differences in reported earnings by type of respondent – observations with usual working hours

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.131*** (0.001)	-0.071*** (0.001)	-0.081*** (0.001)	0.479*** (0.018)	-0.175*** (0.001)	-0.017*** (0.001)	-0.024*** (0.001)	0.273*** (0.019)
Observations	523,449	429,631	429,631	523,449	402,990	326,043	326,043	402,990
Individuals	128,259	124,600	124,600	128,259	96,182	93,737	93,737	96,182
R ²	0.471	0.793	0.799	0.577	0.461	0.820	0.831	0.632
Dep. var. mean (self)	0.900	9.043	3.840	45.806	0.911	8.841	3.688	43.671
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. We also exclude observations in which the worker is not working his/her usual hours. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.12 Differences in reported earnings by type of respondent – workers employed in the same company across survey waves

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
<i>Panel A: Same company according to worker or proxy</i>								
Proxy respondent	-0.129*** (0.002)	-0.073*** (0.002)	-0.083*** (0.002)	0.442*** (0.026)	-0.169*** (0.002)	-0.019*** (0.002)	-0.025*** (0.002)	0.210*** (0.027)
Observations	338,686	278,815	278,815	338,686	261,279	212,880	212,880	261,279
Individuals	105,899	100,491	100,491	105,899	81,285	77,323	77,323	81,285
R ²	0.523	0.819	0.821	0.591	0.513	0.842	0.850	0.652
Dep. var. mean (self)	0.901	9.066	3.875	45.354	0.914	8.862	3.720	43.268
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
<i>Panel B: Same company according to worker</i>								
Proxy respondent	-0.124*** (0.003)	-0.074*** (0.003)	-0.083*** (0.004)	0.422*** (0.051)	-0.157*** (0.003)	-0.019*** (0.003)	-0.023*** (0.004)	0.189*** (0.047)
Observations	148,187	126,849	126,849	148,187	130,299	111,752	111,752	130,299
Individuals	54,545	52,244	52,244	54,545	46,059	44,463	44,463	46,059
R ²	0.536	0.837	0.839	0.632	0.515	0.858	0.864	0.684
Dep. var. mean (self)	0.908	9.072	3.883	45.243	0.919	8.865	3.725	43.174
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. We also restrict our sample to individuals who work for the same company in the period between their first interview and their last interview with the extended questionnaire. In Panel A, we identify workers who remain in the same company using proxy- and self-reported responses to this question. In Panel B, we identify workers who remain in the same company using only self-reported responses to this question. We drop interviews conducted after the last interview with the extended questionnaire. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.13 Differences in reported earnings by type of respondent – workers satisfying all sample restrictions

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.137*** (0.004)	-0.073*** (0.004)	-0.082*** (0.004)	0.437*** (0.049)	-0.179*** (0.004)	-0.024*** (0.004)	-0.030*** (0.004)	0.205*** (0.048)
Observations	87,213	70,071	70,071	87,213	75,261	59,147	59,147	75,261
Individuals	31,287	28,734	28,734	31,287	26,811	24,603	24,603	26,811
R ²	0.566	0.840	0.848	0.689	0.556	0.848	0.862	0.725
Dep. var. mean (self)	0.888	9.158	3.991	44.324	0.891	9.004	3.890	42.073
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The sample is limited to full-time workers (30 or more hours worked per week), formal workers, defined as those who receive access to social security from their employers in all periods, workers who have a salaried position for which they only receive a salary in all periods, interviews not conducted in December, observations for which the reported hours correspond to the number of hours the worker habitually works, and workers who stay at the same firm throughout survey waves. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.14 Differences in reported earnings by type of respondent – controlling for survey round fixed effects

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
Proxy respondent	-0.131*** (0.001)	-0.072*** (0.001)	-0.089*** (0.008)	0.472*** (0.018)	-0.176*** (0.001)	-0.017*** (0.001)	-0.031*** (0.009)	0.255*** (0.019)
Observations	561,300	460,329	33,658	561,300	423,693	342,378	28,385	423,693
Individuals	128,415	125,412	19,497	128,415	96,218	94,208	16,728	96,218
R ²	0.462	0.788	0.893	0.524	0.453	0.816	0.907	0.595
Dep. var. mean (self)	0.899	9.043	3.983	45.559	0.911	8.843	3.876	43.484
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Round FE	✓	✓		✓	✓	✓		✓

Note: The sample is composed of salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressor is a dummy variable that takes a value of 1 if the informant is another household member, and 0 if the worker self-reports. We control for individual and survey-wave fixed effects, as well as survey round fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.15 Relationship between proxy informant and worker

	All	Male	Female
<i>Panel A: Full sample</i>			
Spouse	0.522	0.658	0.323
Child	0.267	0.179	0.396
Parent	0.064	0.046	0.089
Sibling	0.056	0.040	0.078
Other relative	0.062	0.053	0.074
Non-relative	0.014	0.011	0.019
Unknown	0.016	0.013	0.021
Observations	577,883	344,123	233,760
<i>Panel B: Married sample</i>			
Spouse	0.790	0.838	0.673
Child	0.171	0.131	0.268
Parent	0.007	0.006	0.009
Sibling	0.003	0.003	0.002
Other relative	0.028	0.020	0.045
Non-relative	0.001	0.001	0.002
Unknown	0.001	0.001	0.001
Observations	376,658	267,051	109,607
<i>Panel C: Single sample</i>			
Child	0.475	0.392	0.520
Parent	0.187	0.220	0.170
Sibling	0.174	0.207	0.156
Other relative	0.084	0.088	0.082
Non-relative	0.040	0.049	0.036
Unknown	0.039	0.044	0.037
Observations	164,205	57,053	107,152

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. In Panel B, the sample is restricted to workers who are married or in a union and cohabit with their spouse in all periods. The married sample excludes partnered workers for whom we cannot identify the spouse. In Panel C, the sample is restricted to workers that are single in all periods. This table shows the relationship between the worker and his/her proxy for all observations in which the informant is a proxy.

TABLE A.16 Differences in reported earnings by type of respondent – by marital status and relationship with informant

	Male				Female			
	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked	Reports earnings	Monthly earnings (in ln)	Hourly wage (in ln)	Hours worked
<i>Panel A: Married sample</i>								
Spouse informant	-0.078*** (0.001)	-0.073*** (0.001)	-0.084*** (0.001)	0.496*** (0.021)	-0.091*** (0.002)	-0.007*** (0.002)	-0.011*** (0.002)	0.149*** (0.030)
Offspring informant	-0.247*** (0.003)	-0.043*** (0.003)	-0.050*** (0.003)	0.313*** (0.044)	-0.216*** (0.004)	0.002 (0.003)	-0.004 (0.003)	0.256*** (0.046)
Other informant	-0.315*** (0.007)	-0.084*** (0.007)	-0.097*** (0.007)	0.460*** (0.090)	-0.303*** (0.008)	-0.033*** (0.007)	-0.042*** (0.008)	0.429*** (0.099)
Observations	436,583	365,714	365,714	436,583	208,163	174,347	174,347	208,163
Individuals	98,581	96,441	96,441	98,581	47,121	46,316	46,316	47,121
R ²	0.471	0.790	0.793	0.530	0.450	0.821	0.829	0.618
Dep. var. mean (self)	0.900	9.063	3.868	45.496	0.912	8.887	3.751	42.968
<i>Panel B: Single sample</i>								
Offspring informant	-0.237*** (0.005)	-0.091*** (0.005)	-0.104*** (0.005)	0.549*** (0.065)	-0.216*** (0.003)	-0.029*** (0.003)	-0.037*** (0.003)	0.288*** (0.038)
Parent informant	-0.235*** (0.006)	-0.122*** (0.006)	-0.137*** (0.007)	0.592*** (0.086)	-0.222*** (0.005)	-0.067*** (0.005)	-0.080*** (0.005)	0.473*** (0.065)
Sibling informant	-0.229*** (0.006)	-0.056*** (0.006)	-0.066*** (0.006)	0.430*** (0.083)	-0.220*** (0.005)	-0.020*** (0.005)	-0.029*** (0.005)	0.328*** (0.065)
Other informant	-0.245*** (0.008)	-0.052*** (0.008)	-0.062*** (0.008)	0.513*** (0.107)	-0.227*** (0.006)	-0.019*** (0.005)	-0.023*** (0.006)	0.166** (0.081)
Observations	89,538	66,691	66,691	89,538	181,023	140,949	140,949	181,023
Individuals	21,713	20,975	20,975	21,713	41,990	40,883	40,883	41,990
R ²	0.475	0.779	0.781	0.507	0.467	0.809	0.815	0.569
Dep. var. mean (self)	0.891	8.977	3.781	45.566	0.907	8.796	3.634	44.096
Individual FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. In Panel A, the sample is restricted to workers who are married or in a union and cohabit with their spouse in all periods. The married sample excludes partnered workers for whom we cannot identify the spouse. In Panel B, the sample is restricted to workers that are single in all periods. The regressions in columns 1–4 are restricted to male workers, and those in columns 5–8 to female workers. The dependent variable in columns 1 and 5 is a dummy for whether earnings are reported. The dependent variable in columns 2 and 6 (3 and 7) is the monthly (hourly) earnings, both in MEX\$ of 2019 and in ln. The dependent variable in columns 4 and 8 is the weekly hours worked. The main regressors in Panel A are dummy variables for whether the informant is the worker's spouse, child, or another household member (the omitted category is self-reporting by the worker). The main regressors in Panel B are dummy variables for whether the informant is the worker's child, parent, sibling, or another household member (the omitted category is self-reporting by workers). We control for individual and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.17 Gender of Interviewer

	Male worker	Female worker	Total
Female interviewer	0.42545	0.43959	0.43155
Male interviewer	0.57447	0.56034	0.56837
Interviewer gender missing	0.00008	0.00007	0.00008
Observations		926,757	

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The data spans the 2006–2019 period.

TABLE A.18 Gender wage gap – Alternative specifications

	(1)	(2)	(3)
Female	-0.158*** (0.003)	-0.160*** (0.002)	-0.187*** (0.002)
Female × Proxy respondent	0.069*** (0.003)	0.051*** (0.002)	0.052*** (0.002)
Observations	802,707	802,319	802,319
R ²	0.033	0.291	0.443
Time FE	✓	✓	✓
State and urban/rural	✓	✓	✓
Age and age squared		✓	✓
Education		✓	✓
Industry			✓
Occupation			✓

Note: The sample is composed of all salaried workers aged 25–64 who have at least one observation that is proxy-reported and at least one observation that is self-reported. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The dependent variable in all regressions is the worker's hourly wage (in MEX\$ of 2019 and in ln). Our main regressor is a dummy for whether the worker is female, by itself and interacted with an indicator variable for whether the observation is proxy reported. We also include the proxy indicator by itself. In column 1, we also control for survey-wave fixed effects, age, and age squared. In column 2, we additionally control for whether the worker lives in an urban area, state fixed effects, and educational attainment dummies. In column 3, we also control for industry and occupation. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A.19 Gender wage gap – Sensitivity to imputing proxy reports

	Without imputation	Imputing proxy reports
Female	-0.116*** (0.001)	-0.165*** (0.001)
Observations	1,400,888	1,400,888
R ²	0.283	0.330
Controls	✓	✓
Time FE	✓	✓

Note: The sample is composed of all salaried workers aged 25–64. We exclude workers who do not have a salaried position in all periods, and those whose weekly hours do not lie between the 10th and 90th percentiles (30 and 60 hours) of the salaried worker distribution in every wave. The dependent variable in all regressions is the worker's hourly wage (in MEX\$ of 2019 and in ln). In column 2, we impute proxy reports of the dependent variable. For workers who have both proxy and self reported wage, we use the average of self reports. For workers observed only through proxy reports, we impute hourly wages using the average wage of observationally similar workers in the same quarter, matched on gender, state, urban/rural status, educational attainment and age group, and if possible, by industry, and occupation. Our main regressor is a dummy for whether the worker is female. We also control for age, age squared, educational attainment dummies, a dummy for whether the worker lives in an urban area, state fixed effects, and survey-wave fixed effects. Standard errors clustered at the individual level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.